Ancient Roles of Non-coding RNAs in Eukaryotic Evolution

Contributions from Social Amoebae and Other Protists

Lotta Avesson
Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Biology
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2011

Acta Universitatis agriculturae Sueciae
2011:70

Cover: *Dictyostelium discoideum* popart
(photo and design : L. Avesson)

ISSN 1652-6880
ISBN 978-91-576-7614-6
© 2011 Lotta Avesson, Uppsala
Print: SLU Service/Repro, Uppsala 2011

Ancient Roles of Non-coding RNAs in Eukaryotes. Contributions from Social Amoebae and Other Protists.

Abstract

RNAs not coding for proteins, non-coding RNAs (ncRNAs) have many important roles in all kingdoms of life. Especially in eukaryotes, the regulatory functions of ncRNAs have been suggested as a major force in the evolution of complex traits. Cellular processes that are regulated by ncRNAs include for example cell differentiation, organ development and defense against viruses and transposable elements. This is achieved through a number of mechanisms like RNA destabilization and modification, transcriptional and translational control and chromatin modifications.

Dictyostelium discoideum is a social amoeba and the best studied organism representing Amoebozoa, one of the eukaryotic supergroups. It has for long served as an excellent model for many basic cellular events like chemotaxis, differentiation and development and recently also for infection. The ncRNA population in *D. discoideum* is in many ways typical of eukaryotes but also harbors particularities. In this thesis I have studied spliceosomal RNAs as well as the RNA interference and microRNA pathways, which probably were present in the last eukaryotic common ancestor. I have also characterized Class I RNAs which seems to be specific to social amoebae. In addition, we have described the signal recognition particle RNA in several protists and also the involvement of a ncRNA during host interaction and stress in *Giardia lamblia*.

Combining the well established molecular tools and knowledge about various pathways in *D. discoideum*, with the growing understanding of ncRNA, could in the future give important information about the function of ncRNAs as well as their ancient roles and evolution.

Keywords: Dictyostelium discoideum, non-coding RNA, Class I RNA, microRNA, RNA interference, spliceosomal RNA, SRP RNA, evolution

Author's address: Lotta Avesson, SLU, Department of Molecular Biology,
Box 590, 751 24 Uppsala, Sweden
E-mail: Lotta.Avesson@slu.se

Till Farmor, Morfar och Mormor

"If we knew what it was we were doing, it would not be called research, would it?"

Albert Einstein

Contents

List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I *Hinas, A., *Larsson, P., Avesson, L., Kirsebom, L. A., Virtanen, A., and Söderbom, F. (2006) Identification of the major spliceosomal RNAs in *Dictyostelium discoideum* reveals developmentally regulated U2 variants and polyadenylated snRNAs. *Eukaryotic Cell* 5(6), 924-34.

II Ringqvist, E., Avesson, L., Söderbom F., and Svärd, S.G. (2011) Transcriptional changes in Giardia during host-parasite interactions. *Int. J. Parasitol.* 41(3-4), 277-85.

III Avesson, L., Schumacher, H.T., Romby, P., Fechter, P., Hellman, U., Söderbom, F. (2011) An abundant class of non-coding RNA regulates development in the social amoeba *Dictyostelium discoideum*. *RNA Biology* 8(6), *In press*

IV Avesson, L., Reimegård, J., Wagner, E.G.H., and Söderbom, F. Abundantly expressed class of non-coding RNA conserved throughout evolution of Dictyostelia. *Manuscript*

V *Avesson, L., *Reimegård, J., Wagner, E.G.H., and Söderbom, F. The small RNA population in *D. discoideum* during growth and development. *Manuscript*

* These authors contributed equally

Papers I-III are reproduced with the permission of the publishers.

The contribution of Lotta Avesson to the papers included in this thesis was as follows:

I Performed and evaluated some experiments, contributed to the writing.

II Planned, performed and evaluated small RNA experiments including writing and figure preparation.

III Planned, performed and evaluated the majority of experiments. Responsible for the disposition and writing of the manuscript including preparation of all figures.

IV Planned, performed and evaluated all experiments and parts of the bioinformatics. Responsible for the disposition and writing of the manuscript including preparation of all figures.

V Planned, performed and evaluated the majority of experiments. Participated in the design of bioinformatic analysis with J.R.. Responsible for the disposition and writing the majority of the manuscript including preparation of figures.

Abbreviations

aa	amino acid
Ago	Argonaute
cAMP	cyclic adenosine monophosphate
CIBP	Class I binding protein
dsRNA	double stranded RNA
EMSA	electrophoretic mobility shift assay
ER	endoplasmatic reticulum
LECA	last eukaryotic common ancestor
lncRNA	long non-coding RNA
LUCA	last universal common ancestor
miRNA	microRNAs
mRNA	messenger RNA
ncRNA	non-coding RNA
nt	nucleotide
ORF	open reading frame
RACE	rapid amplification of cDNA ends
RdRP	RNA dependent RNA polymerase
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNP	ribonucleoprotein
rRNA	ribosomal RNA
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SRP	signal recognition particle
ssRNA	single stranded RNA
TE	transposable element
tRNA	transfer RNA
UTR	untranslated region

1 Introduction

“Did they teach you what a microRNA is?” was one of the first questions my enthusiastic supervisor asked me when I first came to the lab in September 2004 for a five week project.

“Not really.” I had to admit. So he explained and I was soon very intrigued by this world of “secret” molecules that for some reason was not really mentioned by any textbooks (or teacher or lecturer). I do not know if kids today still learn that the human genome consist of massive amounts of “junk DNA”, DNA that do not encode proteins and therefore have no function. Maybe there are more important things to be right about. In any case, I am very glad I was enlightened and have had the privilege to follow this exciting field of molecular biology from the first row, maybe not from birth but through its adolescence.

1.1 The complex eukaryotic genome

The first studies of genes and gene expression were conducted in bacteria and these data have naturally been the base for our understanding of genome function. This is also when a very essential hypothesis was born, *the central dogma* coined by Francis Crick in 1958 (Crick, 1970), stating that genetic information flows from DNA to RNA to protein (Crick later regretted the use of the word dogma as it caused him quite some trouble). This was for long interpreted as if all genetic information flows from DNA to protein via messenger RNA (mRNA) and that proteins are carrying out not only the catalytic and structural functions in the cell but also most regulatory (Mattick, 2004).

Today, many prokaryotic genomes are sequenced and for them the dogma essentially holds, most of these genomes consist of protein-coding genes. But when various eukaryotic genomes were analyzed some confusion arose. It was soon clear that only a very small part of the human genome encoded proteins

and the term “junk DNA” was born (Ohno, 1972). Equally strange was the fact that the number of protein-coding genes was almost the same in humans, which have approximately 22 000 genes and about 100 trillion cells of maybe thousands of cell types, as in for example the nematode *Caenorhabditis elegans,* which have 19 000 genes and only nearly 1000 cells of about 20 cell types (Arendt, 2008). Furthermore, most of these genes are orthologous, meaning that the most simple animal have a very similar protein toolkit to humans (Rokas, 2008). The number of genes is not proportional to genome size and genome size or gene number cannot be correlated to the (at least intuitive) complexity of the organism (here I consider human as more complex than *C. elegans* due to larger number, and types, of cells). However, it has been suggested, that increasing biological complexity is positively correlated to the ratio between non-coding and protein-coding DNA (after correcting for ploidy) (Figure 1) and this is the only variable that has been shown to do so as of today (Mattick, 2011; Mattick, 2004). This is of course depending on how complexity is measured and have been criticized (Poole, 2004). These kinds of relationships might also be biased by which genomes that have been sequenced to date and it will be interesting to see if they hold when more genome projects are completed.

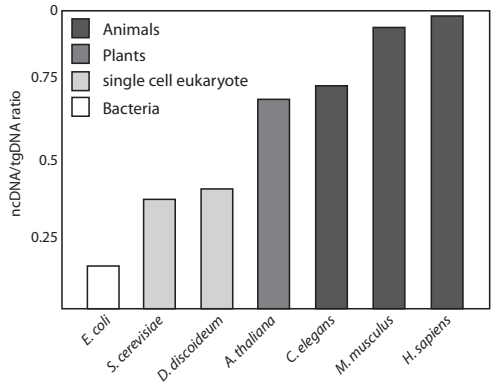

Figure 1. “Organism Complexity” vs non-coding portion of different genomes. ncDNA: non-coding DNA, tgDNA: total genomic DNA. Graph adapted from (Taft *et al.*, 2007).

In any case, during the last decades a growing interest has been directed towards the non-coding genome and it is now clear that eukaryotic genomes are very complex with many layers of regulation that we are far from understanding completely (Costa, 2010; Sharp, 2009). The non-coding DNA has for long been ascribed an important role in the evolution of complex traits

by regulating transcription via *cis*-regulatory sequences (like enhancers and insulators) (Levine & Tjian, 2003). Recently however it has become clear that large portions of the non-coding genome are actually transcribed into RNA. The largest transcriptomics project so far is the ENCODE project with the goal to annotate all elements of the human genome (The ENCODE Project Consortium, 2004). The results from a pilot project representing 1% of the genome was published in 2007 and revealed that up to 90% of the DNA may be transcribed (Birney *et al.*, 2007). Naturally, the next question is of course how much of this RNA is actually functional?

RNA was for long considered simply as an intermediate molecule carrying information from genes to proteins (mRNA) or as an infrastructural component of complexes involved in protein synthesis, for example ribosomal RNA (rRNA), transfer RNA (tRNA) or small nuclear RNA (snRNA). When it became clear that RNA with none of these functions (for example in the form of introns (Williamson, 1977)) exist in the cell, they were generally considered as junk. Later it was acknowledged that these RNA molecules might possess regulatory functions. This was believed to be achieved mainly in *cis* by acting as binding sites for regulatory proteins. Today we know that ncRNAs also can act in *trans*, through many different mechanisms (Mattick, 2009b; Wilusz *et al.*, 2009).

1.2 Classification of life on earth

The classification of life on earth is a matter of ongoing and probably never ending work and debate. For natural reasons, morphology was for long used to determine relationships between different species. But since the emergence of DNA sequencing in the 1950s and development of rapid methods like Sanger sequencing in the 1970s (Sanger *et al.*, 1977; Sanger & Coulson, 1975), molecular phylogeny has taken over. This has certainly caused a revolution in the field. For example, until the 1970s, life on earth was divided into eukaryotes (cells with nucleus) and prokaryotes (cells without a nucleus). But by studying the 16S rRNA genes it was revealed that there were actually two very different groups of prokaryotes and today cellular life is divided into three domains, Bacteria, Archea and Eukarya (Woese & Fox, 1977).

How to further classify eukaryotes is definitely a question of controversy and new analyses are presented continuously. In one hypothesis six eukaryotic supergroups are suggested (Figure 2) (Roger & Simpson, 2009; Simpson & Roger, 2004). There is however still many uncertainties especially regarding the deepest branches, but in this thesis I will refer to the six supergroup classification.

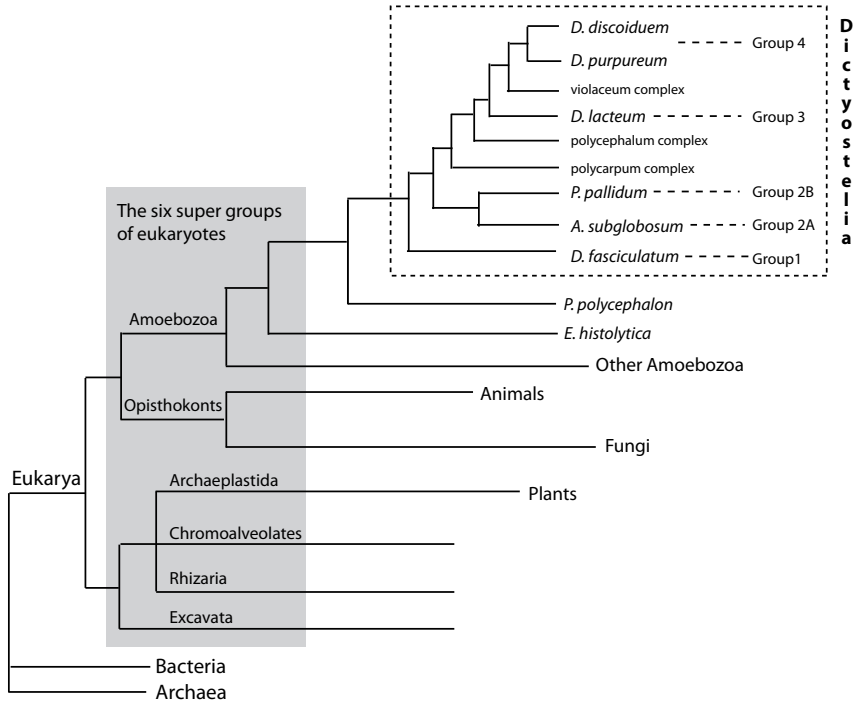

Figure 2. Simplified representation of the evolutionary tree of the eukaryotes, showing the six eukaryotic supergroups and with emphasis on the phylogeny of Amoebozoa and Dictyostelia (adapted from (Roger & Simpson, 2009) and (Romeralo *et al.*, 2011)). Branch lengths are not to scale.

1.3 Amoebozoa and Dictyostelia

As this thesis largely is based on work using the social amoeba *Dictyostelium discoideum* as a model system, I will give an introduction to the Amoebozoa supergroup and Dictyostelia. Several studies have shown that the Amoebozoa supergroup separated from the Opisthokonts before the split between animals and fungi but after the division from the Archaeplastida which includes plants (Figure 2) (Burki *et al.*, 2008; Eichinger *et al.*, 2005; Bapteste *et al.*, 2002). Dictyostelia also called social amoebae are particular in especially one aspect; they can achieve multicellularity by aggregation and cycle between uni- and multicellular stages. The multicellular fruiting bodies are formed by differentiated spore and stalk cells. Dictyostelia have traditionally been classified according to fruiting body morphology into three taxa. *Dictyostelium* is the largest group, containing species with unbranched or laterally branched fruiting bodies. *Polysphondylium* include species with complex multiheaded

fruiting bodies with repetitive whirls of side branches. Species belonging to the third group, *Acytostelium*, form fruiting bodies with acellular stalks.

The first molecular phylogenetic analysis including more than 100 isolates of Dictyostelia, based on small subunit rRNA and tubulin datasets, was presented in 2006 (Schaap *et al.*, 2006). This analysis divided Dictyostelia into four major groups (Figure 2 and 3). Group 1 is a morphologically diverse group of *Dictyostelium* species. Group 2 is a mix of species from all three taxa, including all of the *Acytostelium* species. Also group 3 consists of only *Dictyostelium* species. *D. discoideum* belongs to group 4 which is the largest group. In a recent expanded phylogeny including 50 new species, a division into eight groups was suggested (Figure 2) (Romeralo *et al.*, 2011). Group 2 is split into two groups, 2A consisting of all but one acytostelids and 2B which is the most heterogeneous group. In addition, three complexes containing fairly few species but representing deep and ancient lineages is suggested as major groups. None of the groups defined by molecular phylogeny correspond to the traditional classification and it is obvious that fruiting body morphology is not a good marker for evolutionary relationships among Dictyostelia.

Phylogenetic analyses based on both rDNA and protein-coding genes have revealed the great molecular depth among Dictyostelia, it is equal to that of all animals and much greater than in fungi. In the latest phylogenetic analysis, based on 33 protein-coding genes, the split of the four major groups is estimated to about 600 million years ago (Heidel *et al.*, 2011).

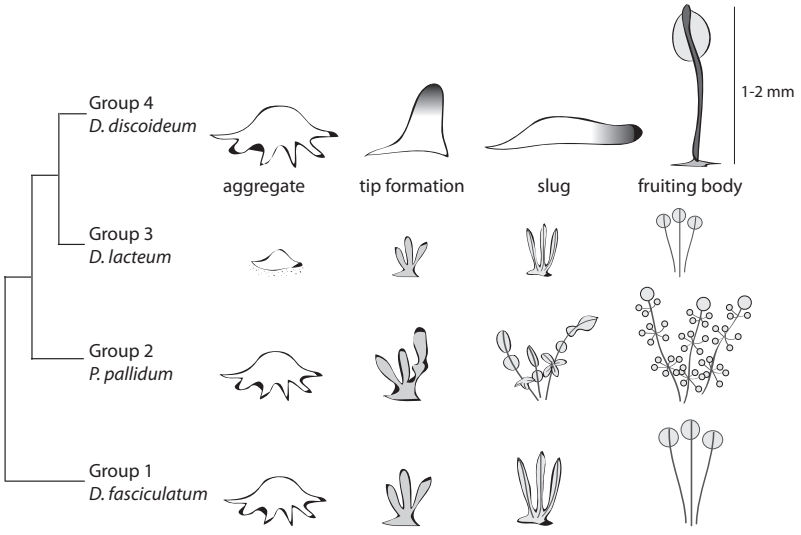

Figure 3. Examples of the morphology of four Dictyostelia species representing different phylogenetic groups. The relative sizes are approximations.

1.4 Dictyostelium discoideum

The slime mold or social amoeba *Dictyostelium discoideum* lives as a haploid single cell amoeba in the soil of the forest floor feeding on bacteria and reproducing by mitotic cell division. The social phase of its life cycle is entered when starvation cause the cells to aggregate. Aggregation is mediated by cAMP signaling. Release of cAMP leads to chemotactic movement and up to 100 000 cells come together to form a mound (Kessin, 2001). A massive change in gene expression is also initiated and at this stage. There are two main fates for the individual cells, either to become a pre-spore or a pre-stalk cell. The pre-stalk cells will be vacuolized and form the stalk in the final stage of the developmental cycle, the fruiting body. The pre-spore cells will form the ball of spores on top of the stalk (Figure 4). In contrast to the dead stalk cells, the spores can germinate and enter the single cell phase again when food is available. The cells commit to development between 4 and 6 h after starvation is induced and do not return to growth phase after this time (Katoh *et al.*, 2007). Before development is complete, during the so called slug stage, Dictyostelium can move around in response to light to search for a good place to culminate. It is remarkable how this transition from free living amoeba to a multicellular organism, where cells communicate and cooperate, can take place in just 12 hours. This social behavior differs from "true" multicellularity in that cell division and differentiation are separated, cells can only feed and divide during the single cell phase.

What are the reasons for such social behavior? The multicellular development has several ecological benefits (Li & Purugganan, 2011). The migrating slug can move further and cross obstacles to find new food sources much easier than the single cell amoeba. The multicellular aggregates also provide protection against predators such as nematodes which are reported to only feed on solitary amoeba. Dispersal of the spores is aided when lifted from the ground by the stalk and spores may also be preserved better when separated from decomposing agents on the ground (Schaap, 2007). But the benefits are not obvious to all individual cells as approximately 20% die in order to form the stalk during development. According to kin selection theory this unselfish behavior can be explained by the high genetic similarity among the individuals in a fruiting body and this has been reported to occur in nature (Gilbert *et al.*, 2007). Whether the high genetic relatedness is maintained only because of passive reasons e.g. close physical distance, or if cells somehow can measure and react to genetic distance is not clear (Li & Purugganan, 2011).

Cheaters are a part of reality in all social cooperation. Cheaters are individuals that use the advantages of the system but do not pay the cost. In the case of *D. discoideum,* cheater genotypes do not contribute equally in the

formation of the stalk but instead form a larger proportion of spores. Many genes have been associated with cheating behavior and cheating mutants are believed to evolve in nature (Santorelli *et al.*, 2008). One example is the *fbxA* mutant which has deregulated levels of intracellular cAMP which is essential during development (Ennis *et al.*, 2000). The fitness of some of these cheater mutants is dependent on frequency and they need to coexist with wild-type genotypes in a balanced way. This is the case for *fbxA* mutants that cannot form fruiting bodies in clonal cultures.

Additionally, *D. discoideum* has a sexual cycle where two haploid cells of different mating types form a diploid zygote. The zygote then attracts surrounding haploid cells by cAMP secretion and cannibalize on them for nutrients. Eventually a dormant macrocyst is formed which later can germinate and release haploid progeny. It is believed that mating and recombination is common in the wild but it has been difficult to obtain in the laboratory (Flowers *et al.*, 2010). *D. discoideum* has three sexes or mating types that all can mate with each other. In a recent study a single genetic locus that specifies the different mating types were identified. The key genes encode two short polypeptides without homology to any known proteins (Bloomfield *et al.*, 2010).

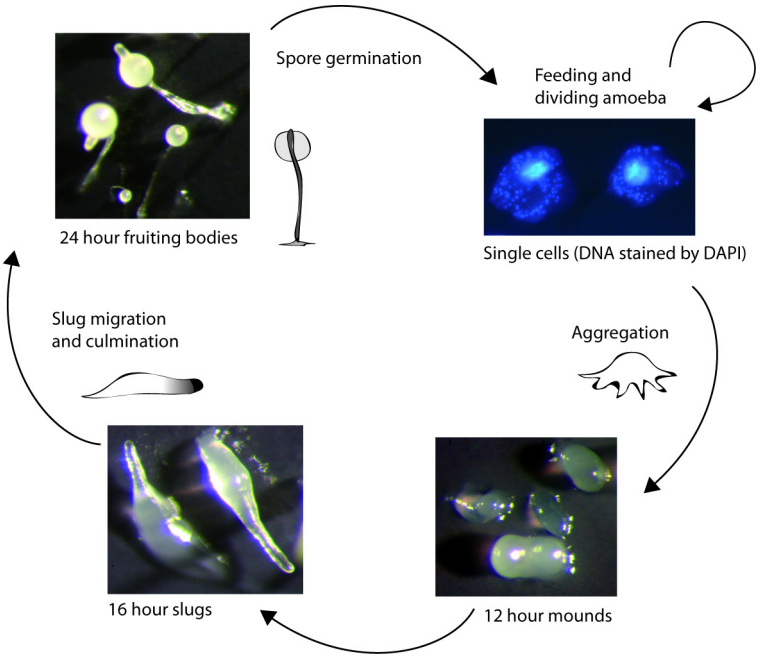

Figure 4. Multicellular development in *D. discoideum*. Photos by L. Avesson

1.4.1 The model organism

D. discoideum was discovered in 1935. It was initially paid interest due to its multicellular development and has served as a very important model for the molecular basis of chemotaxis, cell motility and cell differentiation. *D. discoideum* is also an excellent model for social evolution which includes events like cheating behavior and kin discrimination (Williams, 2010; King & Insall, 2009). In recent years *D. discoideum* has been used to study the mechanisms of infection, in particular that of *Legionella pneumophila*, the causative agent of Legionnaires' disease. *L. pneumophila* naturally infects amoebozoans by avoiding digestion after phagocytosis. *D. discoideum* is a very effective phagocyte and can serve as a model both for natural protozoan host and the human macrophage (Steinert & Heuner, 2005).

The *D. discoideum* cell is generally slightly smaller than an animal cell, it has a flexible plasma membrane which makes it motile and active in pinocytosis and phagocytosis (Kessin, 2001). In the laboratory cells can be grown to high densities in liquid media where they have a generation time of 8-12 hours. The multicellular development is easily induced and synchronized in the laboratory. Among the many molecular tools that can be applied to *D. discoideum* are multiple gene disruption by homologous recombination in combination with the Cre-loxP system, RNAi mediated gene knock-down and random enzyme-mediated insertion (REMI) mutagenesis (Faix *et al.*, 2004; Martens *et al.*, 2002; Kuspa & Loomis, 1992).

1.4.2 The *D. discoideum* genome

The haploid *D. discoideum* genome is 34 Mbp and the sequencing was completed in 2005. The estimated number of protein-coding genes is about 13 500. The overall GC content is very low, only 22% and in some regions of the genome down to a few percent. Most genes contain one or two short introns (Eichinger *et al.*, 2005).

The *D. discoideum* genome contains many repetitive elements including different LTR and non LTR transposons, DNA transposons and yet not classified elements (Glockner *et al.*, 2001). DIRS-1 is a retrotransposon which is present in 40 complete copies and many fragments in the genome (Rosen *et al.*, 1983). It contains 4.1 kb internal sequence flanked by inverted terminal repeats. The internal sequence hosts three overlapping open reading frames (ORFs). The DIRS-1 elements are preferentially inserted into other DIRS-1 elements and are almost exclusively situated at one tip of each of the six chromosomes. These regions are made up of about 50% DIRS-1 sequence but also of other transposon derived sequences like DDT (20%) and Skipper (10%) and are suggested to function as centromeres (Glockner & Heidel, 2009).

DIRS-1 has recently been shown to co-localize with the centromeric histone variant (CenH3) which targets the kinetochores of centromeres (Dubin *et al.*, 2010). Each individual centromeric region is at least 170 kb and together they occupy almost 4% of the chromosomes. *D. discoideum* lack canonical eukaryotic telomeres and instead use ribosomal DNA sequences to maintain the ends of the chromosomes (Eichinger *et al.*, 2005). Although *D. discoideum* is not unique in having atypical telomeres, this invention is not common to all Dictyostelia as other species have been shown to have conventional telomeres (Heidel *et al.*, 2011).

1.5 Non-coding RNAs

Non-coding RNAs (ncRNAs) come in all flavors. They are often very loosely grouped based on size, e.g. large, small and short ncRNAs. Small RNA in bacteria can be several hundred nt, but in eukaryotes the term usually refers to RNAs around 20 nt. Since the focus of this thesis is on eukaryotes I will use "small RNAs" to describe RNAs in the range of 20 to 30 nt. The definition of a long non-coding RNA (lncRNA) is also flexible, but the term will here be used to describe RNAs of several kb. The signature of a ncRNA is of course that they do not contain ORFs and will not be translated into a protein but have a function as an RNA. Some belong to classes of ncRNAs with similar structure and/or sequence elements while others are more unique. Certain ncRNAs are well conserved throughout all kingdoms of life, others have only been identified in one phylum or even in isolated species. I will not attempt to describe all ncRNAs or what their roles are here but give a general view of their various functions, especially those of interest for this thesis. I will also leave out the more well known classes like tRNA and rRNA.

1.5.1 A ncRNA plays a central role in the signal recognition particle

The signal recognition particle (SRP) is a ribonucleoprotein complex responsible for the delivery of proteins to cellular membranes. The key components of the SRP pathway are conserved in all domains of life although there are differences in the functional mechanisms and the composition of the complex.

The mammalian SRP consist of the approximately 300 nt long SRP (7SL) RNA and six protein subunits named after their weight: SRP9, SRP14, SRP19, SRP54, SRP68 and SRP72 (Walter & Blobel, 1982; Walter & Blobel, 1980). The assembled SRP can be divided into two domains. The small or "Alu" domain consisting of the Alu part of the SRP RNA and the proteins SRP9/14,

and the large or “S” domain is formed by the central region of the RNA and the remaining four proteins.

The mechanism of SRP mediated protein targeting is best studied in eukaryotes. Here, the S-domain binds to the signal peptide of the targeted protein as it is emerging from the ribosome. This causes a delay in translation through the action of the Alu-domain allowing the ribosome bound SRP to associate with the SRP receptor anchored to the endoplasmic reticulum (ER) membrane. When SRP subsequently dissociates from SR, elongation resumes and the protein is co-translationally translocated across or integrated into the ER membrane (Pool, 2005).

1.5.2 Catalytic ncRNAs in the spliceosome

The major U2-spliceosome is a ribonucleoprotein (RNP) consisting of five small nuclear RNAs (snRNAs) and a large number of proteins. It is responsible for the removal of most introns from mRNA in eukaryotes and is assembled on the pre-mRNA in a stepwise manner (Wahl *et al.*, 2009). The spliceosome probably evolved in Eukarya after the division from Bacteria and Archaea.

There is an extensive interplay between proteins and the snRNAs during assembly and activation of the spliceosome. The active sites appear to be largely made up of RNA and an intramolecular stem loop in U6 has been shown to mediate the positioning of important metal ions. It is however not clear how the chemical catalysis of splicing is carried out (Valadkhan, 2010).

1.5.3 Regulatory non-coding RNAs

A large set of ncRNAs of variable sizes have roles in regulation of gene expression on different levels rather than having structural or catalytic functions in protein synthesis.

Some ncRNAs form complexes with a set of proteins and function as guides that via complementary base pairing can direct the complex to a target nucleic acid. In this way the RNA provides specificity while the protein executes the function of the complex. The same set of proteins can in this manner be used to regulate many targets by using different ncRNAs as guides. Examples of such guide RNAs are small nucleolar RNAs (snoRNAs) which are involved in chemical modification of other RNAs such as rRNA and snRNA (Reichow *et al.*, 2007). Small interfering RNAs (siRNAs) and microRNAs (miRNAs) described below, are other examples of RNAs that function as molecular guides in the cell.

The functions of most lncRNAs have not yet been characterized, but recent efforts have revealed important regulatory roles (Mercer *et al.*, 2009). lncRNAs can originate from all regions of the genome, they can be derived from

intergenic regions, introns or be antisense to other transcripts. Although these transcripts generally lack conservation across diverse species, their functionality is in many cases supported by a number of other observations (Wilusz *et al.*, 2009). Some examples are cell-specific expression, subcellular localization and dynamic expression influenced by differentiation or other factors. The different roles of lncRNAs uncovered so far include for example influencing transcription by chromatin remodeling or inhibition of pol II recruitment, hybridization to sense/antisense transcripts causing RNA degradation or blocking of splice sites and altering activity or localization of proteins and complexes.

One of the most investigated examples of lncRNAs in mammals is Xist and its antisense transcript Tsix. The Xist gene is 17 kb and located in the X-inactivation centre and it is only expressed by one of the X chromosomes in females. Xist recruit Polycomb repressive complex 2 which leads to inactivation of the chromosome (Lee, 2010).

1.5.4 microRNAs and small interfering RNAs

The most significant reason for the incredible rise in interest in ncRNAs during the last decade is the discovery of small (20-30 nt) RNAs. The exploration of these small RNAs in a wide range of organisms has been possible due to the development of high throughput sequencing techniques (see section 1.7.1).

These small RNAs engage the RNA interference (RNAi) related pathways which have been shown to play exceptionally important roles in gene expression and genome defense in eukaryotes (Carthew & Sontheimer, 2009; Ghildiyal & Zamore, 2009; Voinnet, 2009). RNAi is used to describe different gene silencing mechanisms which are induced by RNA. There are three main classes of small RNAs, siRNAs, miRNAs and Piwi-interacting RNAs (piRNAs), involved in RNA silencing. siRNAs are derived from long double stranded RNA precursors which can originate from both exogenous and endogenous sources like viruses and transposable elements. miRNAs are generated from endogenous transcripts that fold into hairpins (Figure 5). These precursors are usually of intergenic origin but introns as well as ncRNAs such as snoRNAs have also been shown to act as miRNA precursors (Ender *et al.*, 2008; Saraiya & Wang, 2008). The third class of small RNAs, piRNAs, have so far only been found in animals where they are involved in regulating transposable elements (TEs) in the germline (Vagin *et al.*, 2006). Since piRNAs do not seem to exist in *D. discoideum* I will not describe their biogenesis or function here.

The RNAi machinery

There are three key players of the RNAi machinery. First, the RNase III-type Dicer proteins, which cleave the double stranded RNA (dsRNA) precursors leaving a 5' monophosphate and two nt 3' overhang as signature (MacRae & Doudna, 2007). Second, the Argonautes (Agos), which is a large family of proteins present in various numbers in different organisms. Some Agos have catalytic function and cleave the target RNA. Others are not catalytic but instead affect mRNA stability or translation through various mechanisms. The combination of the small RNA and a certain Ago determines which RNAs that will be targeted and in what way (Czech & Hannon, 2011). The third major player is RNA dependent RNA polymerases (RdRPs). RdRPs can amplify a silencing signal by producing secondary siRNAs. This can be accomplished by using single stranded RNA (ssRNA) templates either to produce long dsRNA that can be processed by Dicer or by directly making short RNAs *de novo*. RdRPs have been found in most organisms with functional RNAi systems, but canonical RdRPs are missing in for example mammals and insects (Maida & Masutomi, 2011).

Several other proteins associated with the RNAi pathway have been identified in different organisms. One example is GW182 proteins and its homologs which have been shown to be associated with Agos in animals (Meister *et al.*, 2005). GW182 are required for miRNA mediated silencing and act at the effector step, downstream of Agos. No homolog has so far been found outside animals. In plants, the protein Hen1 methylates the 3' end of miRNAs and siRNAs which protect them from degradation (Li *et al.*, 2005). Hen1 orthologs have been shown to methylate piRNAs and siRNAs in animals, for example Drosophila, although the role of this modification is not determined (Horwich *et al.*, 2007).

Biogenesis of siRNAs and miRNAs

The canonical pathway of miRNA biogenesis in plants and animals is described in Figure 5. This is a simplified view and many alternative pathways are utilized in different organisms (Miyoshi *et al.*, 2010). Both siRNAs and miRNAs are processed from their dsRNA precursors into short duplexes (20-25 bp) by Dicer proteins. The duplex is then associated with an Ago and other factors to form the RNA-induced silencing complex (RISC). As one strand of the short RNA duplex (referred to as passenger strand or miRNA*) is removed, the remaining strand can function as a guide through complementary base pairing, bringing RISC to its target RNA (Chapman & Carrington, 2007; Bartel, 2004). The target is then normally downregulated through different mechanisms including RNA destabilization, cleavage or translational inhibition

(see the following section). In this way, miRNAs for example regulate the expression of thousands of mRNAs and thereby basically every cellular process in humans. It has been suggested that at least 60% of all human protein-coding genes are conserved miRNA targets (Friedman *et al.*, 2009).

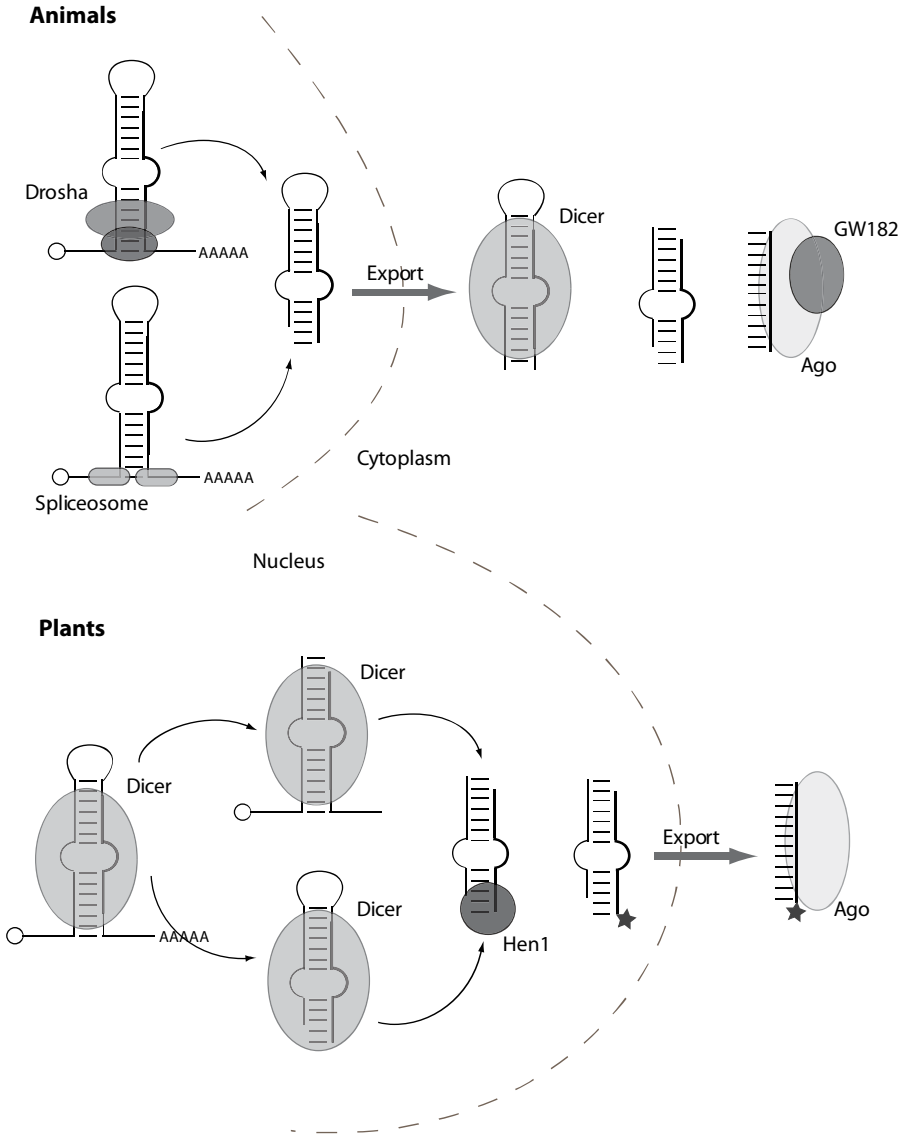

Figure 5. General miRNA pathway in animals and plants.

miRNA target recognition mechanisms

Target recognition and mode of regulation by miRNAs differs in several aspects between different organisms. Some of these are described in Figure 6 (Huntzinger & Izaurralde, 2011; Bartel, 2009). Initially, evidence suggested that miRNAs repress their targets at the level of translation in animals and by inducing cleavage of the mRNA in plants. It is now becoming increasingly clear that the situation is rather complex and far from understood and I will here give a very brief overview of the current models.

Plant miRNAs recognize completely or almost completely complementary target sites, generally located in ORFs. This causes cleavage of the mRNA catalyzed by the Ago protein between nt 10 and 11 opposite the miRNA (Llave *et al.*, 2002; Rhoades *et al.*, 2002). The mRNA fragments are then degraded. It has been reported however that translational repression might be more common in plants than expected (Brodersen *et al.*, 2008).

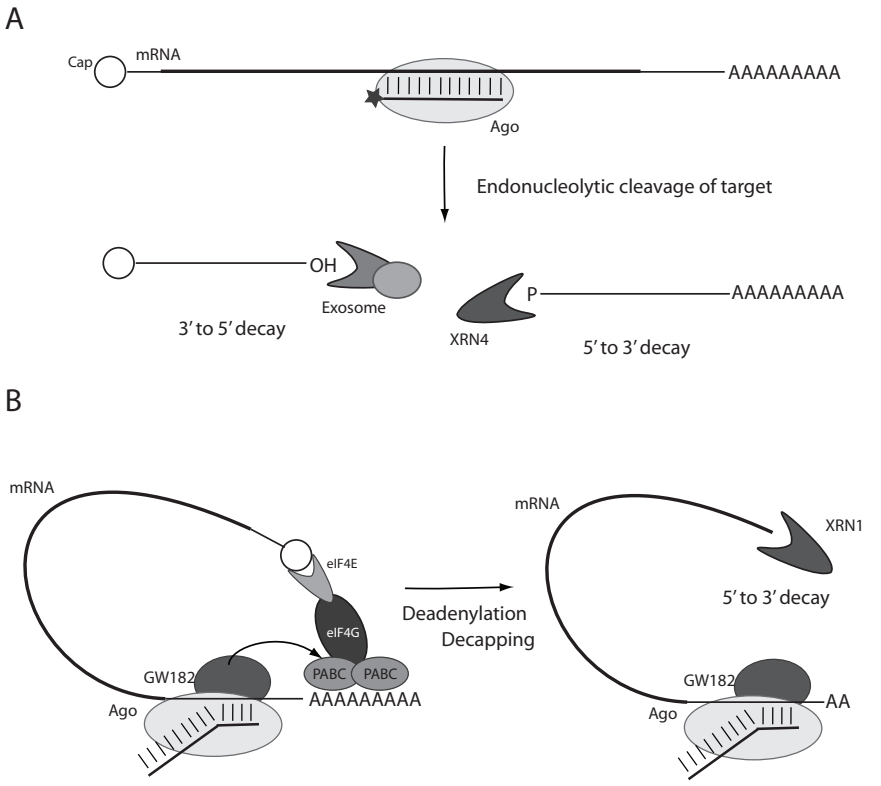

Figure 6. Simplified view of (A) the main mechanism of miRNA mediated silencing in plants and (B) one possible mechanism of miRNA mediated silencing in animals. Adapted from (Huntzinger & Izaurralde, 2011).

In animals, miRNAs normally recognize partially complementary and often multiple target sites in the 3' untranslated region (UTR) of the mRNA. The seed sequence (nt 2-8) of the 5' end of the miRNA is the most important determinant for target recognition (Lewis *et al.*, 2003). A number of mechanisms have been suggested for miRNA silencing in animals. These include repression of translation through several mechanisms, for example block of initiation (Huntzinger & Izaurralde, 2011) and mRNA destabilization via deadenylation (Eulalio *et al.*, 2009) and/or decapping (Eulalio *et al.*, 2007). Recent studies propose that mRNA decay is the major consequence of miRNA regulation in animals (Guo *et al.*, 2010; Baek *et al.*, 2008; Selbach *et al.*, 2008). It is still unclear however if degradation of mRNA targets is a consequence of an initial block of translation or a primary mechanism of silencing.

1.5.5 Other classes of ncRNAs

There are several other classes of small RNAs that have been found in eukaryotes that are worth mentioning. The functions of these RNAs are only partly, at best, elucidated.

Vault particles are enormous complexes (three times the size of the ribosome) found in many eukaryotes. Their function is largely unknown but a role in drug sequestration has been suggested (Izquierdo *et al.*, 1996). In addition to a number of proteins, the vault particle consists of vault RNA (vtRNA). However, there are examples (including *D. discoideum*) where the vault particle but no vtRNAs have been identified. The size of vtRNAs vary from about 80 to 150 nt and the 5' and 3 ' ends can base pair to form a stem. The vtRNA genes are normally present in a single or small number of clustered copies (Stadler *et al.*, 2009).

Y RNAs were discovered already in the 1980s when they were isolated as a component of the Ro RNPs in humans (Lerner *et al.*, 1981). There are four Y RNA genes in mammals and homologs are found so far in vertebrates and in nematodes. Interestingly one homolog is also present in the bacterium *Deincoccus radiodurans* (Chen *et al.*, 2000). These genes are transcribed by pol III into short (about 70-110 nt) stem loops. It has been reported that Y RNAs play a role during DNA replication (Christov *et al.*, 2006). Y RNAs have also been suggested to be a repressor of Ro and Ro RNPs are involved in RNA quality control by binding to misfolded RNAs (O'Brien & Wolin, 1994). Thus, it has been suggested that Y RNAs have several distinct modes of action (Langley *et al.*, 2010).

Another class of ncRNA is stem bulge RNAs (sbRNAs) that were discovered in *C. elegans* (Deng *et al.*, 2006). sbRNAs can also form short stems intervened by variable loop sequences. In an extended search for

sbRNAs in Nematoda it was recently shown that sbRNAs are actually homologs of Y RNAs (Boria *et al.*, 2010). It appears however that sbRNAs might have evolved into a functionally distinct class of ncRNAs since they do not bind to Ro60 ortholog in *C. elegans* (Van Horn *et al.*, 1995).

Of course, the *D. discoideum* Class I RNAs also belong to this group of less characterized RNAs, but these will be discussed in more depth in later sections.

1.6 RNA – evolutionary aspects

RNA is not only playing a central role in modern cells but is also believed to be *the* molecule of prebiotic life. The RNA world theory suggests that the last common ancestor of all life on earth was preceded by life forms based on RNA. According to this theory, RNA functioned both as carrier of genetic information and as a catalyst. The discovery of catalytic RNAs in the 1980s is regarded as the most important support for this theory (Gesteland *et al.*, 2006). The RNA world theory has gained a lot of support over the years but it is controversial and has been extensively criticized (Kurland, 2010). Moreover, another burning question is whether life on earth in that case actually began with RNA (RNA first) or if RNA was preceded by another replicating, evolving molecule (RNA later). There are many difficulties, not least chemical aspects, in explaining either of these views as well as the RNA world theory as such and I will not attempt to go into details here.

In most modern life forms DNA has taken the role as genetic material and proteins are carrying out the majority of the catalytic reactions. However, with the recent discoveries of the various functions of ncRNAs it surely makes sense to talk about a modern RNA world. Still, little is known about the evolution of these pathways and their components.

Many RNAs involved in protein synthesis (rRNA, tRNA, RNase P RNA, SRP RNA) were present in the last universal common ancestor (LUCA) while others, which generally have regulatory functions, have evolved only in certain lineages (Bompfunewerer *et al.*, 2005) (Figure 7).

1.6.1 RNA interference in eukaryotic evolution

It is believed that a RNAi machinery with at least one Dicer, Ago and RdRP was present in the last common eukaryotic ancestor (LECA) (Cerutti & Casas-Mollano, 2006). This machinery was responsible for the production of small RNAs from double stranded precursors and the silencing of cognate sequences, serving as a defense against viruses and various transposable elements. Apparent ancestors of the key protein components (or rather domains) of RNAi

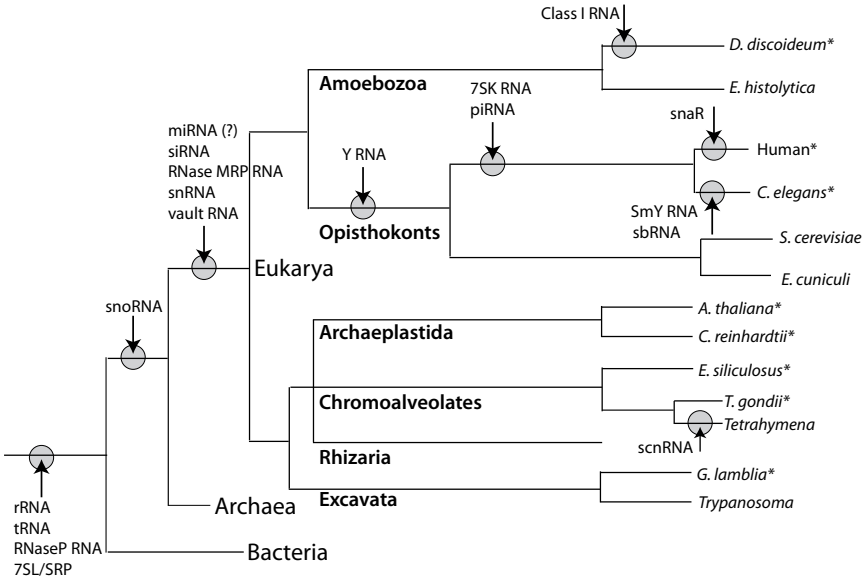

Figure 7. An overview of the emergence of some important ncRNAs in eukaryotes. Asterisk following a species name indicates that miRNAs (or miRNA-like RNAs) have been identified.

have also been identified among prokaryotes although no homologous system has been identified (Shabalina & Koonin, 2008). The RNAi machinery appears to have been lost in many eukaryotes, like for example *S. cerevisiae*, *Trypanosoma cruzi* and *Plasmodium falciparum*. There are also lineages where certain components are missing, for example mammals which lack canonical RdRPs. So what about miRNAs, were they present in LECA? miRNAs were discovered first in animals and then in plants (Reinhart *et al.*, 2002; Lee *et al.*, 1993). At that point it was believed the systems had evolved independently. This was supported by the differences in biogenesis and action of miRNAs between plants and animals and also that no conserved miRNAs were identified across the kingdoms (Jones-Rhoades *et al.*, 2006; Millar & Waterhouse, 2005). Furthermore, no miRNAs could be detected in several single cell organisms indicating that they played a role in multicellularity. In 2007 however, miRNAs were reported both in *D. discoideum* (Hinas *et al.*, 2007) and in the single cell green algae *Chlamydomonas reinhardtii* (Molnar *et al.*, 2007; Zhao *et al.*, 2007). In 2008 miRNA like small RNAs generated from snoRNAs were found in *G. lamblia* (Saraiya & Wang, 2008) and two years later miRNAs were reported in the parasite *Toxoplasma gondii* (Braun *et al.*, 2010). The same year an analysis of the brown algae *Ectocarpus siliculosus*, which is multicellular but belongs to a different supergroup than plants and green algae, revealed a number of miRNAs (Cock *et al.*, 2010). Hence,

miRNAs have now been reported in five of the six eukaryotic supergroups, indicating that some kind of proto-miRNAs might have played a role in eukaryotes early on, before the divergence of the supergroups.

There are several models describing how new miRNAs genes can evolve. In plants, it appears as if some miRNA genes originate from inverted duplication of the target gene. Transposable element and repeat sequences seem to be a common origin of miRNA genes especially in animals (Piriyapongsa *et al.*, 2007; Smalheiser & Torvik, 2005). Genomes furthermore naturally encode many hairpins with the potential to develop into new miRNA genes. Duplication-mutation events of existing miRNA genes are well supported in animals where miRNA genes are often clustered (Bartel, 2004). Also, since a miRNA precursor can potentially produce two functional miRNAs, arm switching seems to be a general mechanism to evolve new functions of a miRNA gene (Griffiths-Jones *et al.*, 2011).

1.7 Finding and annotating ncRNAs

1.7.1 Experimental approaches

During the work with this thesis, high-throughput (deep) sequencing technologies have revolutionized genome sequencing and transcriptomics, not least small RNA discovery. During the first half of the ‘00s, RNA expression was commonly studied by traditional sequencing of cloned cDNA products or microarrays. Today, several next generation platforms like Illumina (Solexa), Roche 454 and Life Technologies (SOLiD) for RNA sequencing are available which all sequence many billions of DNA strands in parallel, generating massive amounts of data (Ozsolak & Milos, 2011).

Although these technologies have been tremendously useful, there are still limitations. The fact that most current RNA sequencing methods rely on cDNA synthesis is limiting for many applications. But also other steps, including fragmentation, ligation and amplification can introduce various errors. RNAs are for example often modified in their 5’ and/or 3’ ends which can inhibit ligation. In particular for small RNAs, quantification has been proven difficult and this seem to depend on small RNA library preparation rather than the sequencing platform (Linsen *et al.*, 2009). Some of these problems might be solved by the use of direct RNA sequencing. The first platform for parallel direct RNA sequencing has been developed by Helicos. In their approach a single molecule of poly(A)-tagged RNA is hybridized to a poly(dT) coated surface and sequenced. One of the main advantages is that all RNA species, long and short, can be analyzed in the same experiment. This is not possible with cDNA based methods since different strategies are required depending on

the size of the RNA. These advances will hopefully diminish the problems of today and allow us to study accurately the transcriptomes of minute cell number and even single cells (Ozsolak *et al.*, 2009).

1.7.2 Computational approaches

Genomes can be searched for known and new gene families using computational approaches. For ncRNA genes this is often more difficult than for protein-coding genes. The nucleotide sequence of protein-coding genes are generally under strong negative selection to maintain the amino acid sequence while ncRNAs are more dependent on structure (Eddy, 2002). Fast evolving sequence and the often short length disqualify commonly used programs like BLAST when searching for ncRNA genes. As many classes of ncRNA fold into specific structures, but also display short conserved sequence elements, the best alternative is many times to include both structural and sequence information. One such tool is the Infernal software, which combines consensus sequence with secondary structure to search DNA sequences (Nawrocki *et al.*, 2009). Presence of promoters and terminators can also be included and used as search criteria. In particular, many algorithms have been developed to identify miRNAs, for example Mirfold and miR-abela. These programs predict if a sequence contain pre-miRNA structures. Mirfold and miR-abela are optimized on plant and animal miRNAs respectively (Boccara *et al.*, 2007; Sewer *et al.*, 2005).

1.7.3 miRNA annotation criteria

Of particular interest for this study has been the criteria for miRNA annotation. The first guidelines came in 2003, these primarily required evidence of expression by for example cloning or northern blot and the correct folding of the surrounding sequence into a miRNA precursor structure (Ambros *et al.*, 2003). However, these criteria have become stricter as the knowledge about miRNA biogenesis and expression has increased. Since the miRNA pathway differs slightly between plants and animals, not the exact same rules can be applied for all organisms. The more complex the population of small RNAs within an organism is, the more difficult it is to distinguish miRNAs from various siRNAs. In general, plants have a more complex pool of endogenous small RNAs than animals due to the presence of PolIV/PolV dependent siRNAs and a high prevalence of secondary siRNAs and in 2008, a new set of criteria was presented for the annotation of plant miRNAs (Meyers *et al.*, 2008). The most fundamental feature of a miRNA is the precise processing of single stranded stem-loop precursor into an approximately 21 nt duplex consisting of the miRNA and miRNA*. This duplex should have a two nt 3’

overhang, indicative of Dicer cleavage, and base pairing should be extensive, typically less than four mismatches. Asymmetric bulges should be infrequent and small throughout the whole stem-loop precursor. Other characteristics can significantly increase the confidence of the annotation but is not necessary or sufficient on their own. These include conservation, identification of target(s), Dicer dependence and independence of RdRPs and PolIV/PolV.

Today deep sequencing naturally is the most common source of expression data used for miRNA searches. This allows for detection of rare miRNAs but increasing sequencing depth also makes it challenging to distinguish true miRNAs from fragments of other transcripts. Some guidelines were recently collected by miRBase to standardize the searches (Kozomara & Griffiths-Jones, 2011). These include for example support from multiple reads, preferably from independent experiments, presence of miRNA* sequences and a defined 5' end.

1.8 Non-coding RNAs in *D. discoideum*

The first ncRNAs in *D. discoideum* were discovered already 20 years ago. In 1992 it was shown that the prespore gene *psvA* was regulated by a 1.8 kb antisense transcript (Nellen *et al.*, 1992). This was the first such example described in a eukaryote. Just two years later dutA, a lncRNAs with mRNA like features was reported (Yoshida *et al.*, 1994). The function of the dutA RNA is not clear but it is developmentally regulated and has a defined expression pattern in different prestalk cell types.

Until recently however, very little was known about the ncRNAs in *D. discoideum*. But experimental and computational approaches during the last seven years have led to many both expected and unexpected findings and at least a basic knowledge about the ncRNA population in this model organism. The starting point was the construction of full-length cDNA libraries representing RNAs from 50 to 500 nt. In this study 17 box C/D and one box H/ACA snoRNAs, SRP RNA and U2 snRNA were identified (Aspegren *et al.*, 2004). Importantly, two novel classes (now redefined as one class (Avesson *et al.*, 2011)) of ncRNAs were also discovered. These are described more in section 1.8.1. Later, the expression of RNase P and RNase MRP RNA was confirmed by us (Avesson unpublished).

1.8.1 Class I RNAs

A large fraction of the 50-150 nt cDNA library represented a group of RNAs without apparent homology to any known classes of ncRNAs. They were named Class I and Class II RNA, are 40-65 nt long and the 5' and 3' ends were

predicted to form short stems. Following the 5' part of the stem is a conserved 11 nt sequence motif (Aspegren *et al.*, 2004). They were further shown to be mainly localized to the cytoplasm and downregulated during development. 14 Class I and two Class II RNAs were represented in the cDNA library and 24 additional genes could be computationally predicted. A similar number of genes were also found in an independent computational search (Mosig *et al.*, 2006). We have now redefined Class I and Class II RNAs as one class (Class I a and b) and the further characterization of these RNAs has been one of the major focuses of this thesis.

1.8.2 The RNAi pathway and miRNAs

Homologs to important RNAi machinery components are present in the *D. discoideum* genome. These include two Dicers, five Agos and three RdRPs (Cerutti & Casas-Mollano, 2006; Martens *et al.*, 2002). RNAi has also been used to silence gene expression in *D. discoideum* (Martens *et al.*, 2002) although this technique is not yet widely used. However, no endogenous RNAi related small RNAs had been identified in *D. discoideum* until 2007 when the first cDNA library of small (18-26 nt) RNAs was published (Hinas *et al.*, 2007). This revealed a very high abundance of 21 nt siRNAs derived in particular from the DIRS-1 retrotransposon but also from other repeat elements. In addition, two Dicer B dependent miRNAs were identified in this study. Although the major players of the RNAi pathway are present in *D. discoideum* very little is known about which complexes that are formed and by what mechanism they are acting.

2 Present investigation

In this thesis I have primarily studied both a new class of ncRNA, Class I RNAs, which was discovered in *D. discoideum* and seems to be specific to Dictyostelia, and also the RNAi/miRNA pathways in *D. discoideum* that probably were present in some form in LECA. I have also been involved in work including the characterization of the spliceosomal RNAs in *D. discoideum* and unusual SRP RNAs from a number of protists as well as small RNAs induced during host interaction and stress in *Giardia lamblia*. Hopefully my work will contribute to the understanding of the functions of ncRNAs in *D. discoideum* but also in a wider perspective, to the role of ncRNAs in eukaryotic evolution.

2.1 Investigation of Class I RNAs (Paper III-IV)

Possibly the most exciting finding in the cDNA library representing 50-150 nt RNAs in *D. discoideum* was the discovery of a novel class of ncRNAs, Class I RNAs (Aspegren *et al.*, 2004). To further understand the function of these RNAs a number of approaches were taken.

2.1.1 Class I RNAs interact with one or several proteins (Paper III)

Since small ncRNAs rarely, if ever, appear naked in the cell, identification of interacting partners is an important step towards understanding their function. First, complex formation was studied using electrophoretic motility shift assay (EMSA) of *in vitro* transcribed Class I RNAs in cell extract. This revealed that Class I RNAs interact with one or several proteins. To further investigate the interactions of endogenous Class I RNAs, cell extracts were separated on sucrose gradients and subsequently fractionated. The RNA content of the different fractions was analyzed by northern blot and it was shown that Class I RNAs are mainly present in the same fractions as SRP RNA but not in the

ribosomal or polysomal fractions. This indicates that Class I RNAs form a relatively small RNP complex and is not associated to any large extent to the ribosome. Hence, a direct role in translational control seems unlikely.

To identify these interacting proteins, *in vitro* transcribed and biotinylated Class I RNA was used in a pull down assay. Several proteins were identified by mass spectroscopy but the most interesting was a 293 aa long protein containing two RNA recognition motifs (RRM). This protein, from here on called CIBP for Class I binding protein, was expressed and purified and shown to bind Class I RNAs in new EMSAs.

2.1.2 *In vitro* probing of Class I RNA structure (Paper III)

The structure of ncRNAs can often be as important, or even more essential for their function, as their primary sequence. Since Class I RNAs are predicted to form a short stem by base pairing of the conserved 5' and 3' ends we wanted to test this by *in vitro* probing. By using various methods, i.e. In line, chemical and enzymatic probing it was shown that the predicted stem indeed is formed *in vitro*. In addition, it seems like the conserved 11 nt sequence following the 5' part of the stem, remains in an open conformation.

2.1.3 Deletion of Class I RNA genes (Paper III)

Disruption of protein-coding genes by homologous recombination is well established in *D. discoideum* (Faix *et al.*, 2004). However, no one had previously reported the disruption of small ncRNA genes. This is potentially problematic since intergenic regions where these genes are located are extremely AT-rich. This could cause the constructs used for recombination to be too unspecific. In addition it is very difficult to design primers to amplify these sequences by PCR. To solve this, deletion constructs were designed with longer than normal homologous regions that overlap with at least one flanking protein-coding gene (which have higher GC content).

Using this design two Class I RNA genes *ddR-21* and *ddR-33*, were successfully deleted. The two deletion strains were subject to phenotypic analysis. Growth rates on bacterial lawns as well as in liquid culture were comparable to wt cells. During development however, a subtle phenotype was observed for the *ddR-21* deletion strain. Compared to wt, these cells formed more but smaller fruiting bodies.

2.1.4 Identification of Class I RNAs in other Dictyostelia (Paper IV)

How ancient are Class I RNAs? Are they specific to *D. discoideum* or can homologs be found in other Dictyostelia, more distant amoebozoans or even in other supergroups? This question was basically impossible to answer until very

recently since only the sequence of the *D. discoideum* genome was available. However, in 2010 the genomes of *Dictyostelium purpureum* (Group 4), *Dictyostelium lacteum* (Group 3), *Polysphondylium pallidum* (Group 2B*)*, *Acytostelium subglobosum* (Group 2A) and *Dictyostelium fasciculatum* (Group 1) were completely or almost sequenced although not assembled (Figure 2). This allowed us to search for Class I RNA genes also in these species. Since the conserved parts of Class I RNAs are very short, there is not much information that can be used for homology searches based purely on sequence. We therefore used the Infernal software (Nawrocki *et al.*, 2009) to perform searches based both on secondary structure and conserved sequence (except for *A. subglobosum*, which was subject to BLAST searches only). The search was combined with manual identification of the upstream element DUSE. Using this method we identified between six and 23 Class I gene candidates in these different genomes. Interestingly, the stem forming 5' and 3' ends differ in sequence between the different species while the 11 nt sequence motif is more conserved. This further supports that the formation of the stem is important but not its sequence and that the 11 nt motif plays an essential role in Class I RNA function.

To investigate if these genes are expressed, RNA was prepared from all five species and analyzed by northern blot. At least two different RNAs were probed in each species and in all cases expression could be easily detected. Infernal searches for Class I RNA genes were also performed on the *Entamoeba histolytica* and *Physarum polycephalum* genomes, the only two additional amoebozoan genomes available. Furthermore, to include a metazoan species, the *C. elegans* genome was searched. Only a few sequences with scores just above the threshold used for the Dictyostelia species was found. Manual inspection disqualified the *E. histolytica* hits as false but for *P. polycephalum* two candidates seemed reasonable although no good upstream element could be found. We also managed to get hold of RNA from *P. polycephalum* but expression of the candidate Class I genes could not be detected by northern blot. It should however be noted that the evolutionary distance between Dictyostelia and these other amoebozoans is very long (Watkins & Gray, 2008; Song *et al.*, 2005) and if Class I RNA are present they may be too divergent to be detected by our search model. One sequence above threshold was also identified for *C. elegans* but no expression could be detected by northern blot.

2.1.5 Phylogeny of Class I genes (Paper IV)

One obvious question is how and when the expansion of Class I RNA genes took place. Several approaches were used to understand the phylogeny of the

genes identified in the different Dictyostelia species. This is however not trivial since Class I RNAs are short and the conserved parts are even shorter. It is thus very difficult to make good overlapping alignments suitable for phylogenetic analysis. One approach is to make alignments based on secondary structure. However, the structure of the region between the short stem connecting the 5' and 3' ends seems to be rather flexible, and also quite variable in length, implying that the overall secondary structure is not suitable for this kind of analysis. In the end, partial and manually edited sequence based alignments representing the 5' and 3' ends plus some of the intervening sequence, were used to construct a phylogenetic tree using maximum likelihood. Considering the low number of informative positions in the alignment, the sequence of the stem will have a very strong influence over the suggested phylogeny. In line with this we could observe that Class I genes from the same species also cluster together in the tree.

Shared synteny or co-localization of genes between species is another way of determining if genes are originating from the same original copy. By analyzing protein-coding genes in the vicinity of Class I genes in the Group 4 species *D. discoideum* and *D. purpureum,* only one obvious syntenic locus was identified. The *D. discoideum* gene is however a rather degenerated pseudogene which lacks the 5' part of the stem completely and does not seem to be expressed. It is thus difficult to make a good comparison of the two genes. In conclusion, this suggests that the expansion of Class I genes largely took place independently in different Dictyostelia.

2.1.6 Additional experiments

Some additional approaches were taken to study the function of Class I RNAs. For example, we attempted *in situ* hybridization to determine the localization of DdR-21 both in single cells and in multicellular developments (slugs). In single cells we failed to obtain a strong hybridization signal but we did detect weak signals from speckles in the cytoplasm. However, these results were too unreliable to draw any conclusions from. When using whole slugs it appeared as if the probe did not penetrate the cells and we did not obtain specific signal. Some effort was also put into characterizing the protein (CIBP) identified in the Class I RNA pull down. This was done by overexpressing FLAG-tagged CIBP and analyzing interacting RNAs after immunoprecipitation. This strategy could unfortunately not be optimized within the timeframe of this thesis. Furthermore, we attempted to knock out CIBP by homologous recombination, but no clones were obtained despite repeated transformations. This could imply that CIBP is essential in *D. discoideum*.

2.2 Deep sequencing of small RNAs in *D. discoideum* (Paper V)

The limited cDNA library of 18-26 nt RNAs from 2007 provided good insight into the small RNA population in *D. discoideum* (Hinas *et al.*, 2007). We expected however, that this library was far from saturated since for example only one clone each of the two miRNAs was present. Therefore, we took advantage of the recent development and availability of deep sequencing methods (in this case the SOLiD platform) to find more rare transcripts. We were also interested in investigating the slightly longer RNAs, since for example piRNAs are approximately 26-30 nt. To be able to study changes in the small RNA population during development, RNA from growing cells (0h), as well as cells from slugs (16 h) and fully developed fruiting bodies (24 h) were sequenced.

2.2.1 The small RNA population

The SOLiD sequencing generated 4-6 million reads per library. After removing adaptor sequences and reads not mapping to the genome, 0.7-1.5 million reads remained. Sequences between 15 and 34 nt were mainly considered for our small RNA analysis. As anticipated the majority of the reads are exactly 21 nt in all libraries.

2.2.2 Identification of miRNAs

In our search for potential miRNAs only 21 nt sequences were considered since the two known miRNAs have this length. First, reads originating from repeat regions present at more than 9 locations in the genome were sorted out. Then, regions of the genome that are transcribed from both strands were removed. This was achieved by only considering locations generating at least 20 times more reads than a region 500 nt upstream and downstream from the opposite strand. Only locations with more than five mapping sequences were considered. After this filtering, between 4000 and 13000 reads mapping to about 250 locations remained in each library. 24 of these locations were predicted to fold into pre-miRNA structures with high confidence by Mirfold. Among these were the two previously identified miRNAs ddi-mir1176 and ddi-mir-1177, but also their predicted miRNA* sequences. Of the remaining locations, six can be regarded as miRNA/miRNA* pairs, leaving us with 17 new miRNAs or 19 miRNAs in total. In addition, miRNA* sequences at levels below five reads were identified for eight miRNAs, but for four of the miRNAs no miRNA* was found.

2.2.3 Origin and nature of miRNAs

The majority of the miRNAs map to unique intergenic regions but some originate from multiple locations in the genome which are derivatives of the repeat element Thug-S. Thug-S resembles the miniature-repeat transposable elements (MITEs) (Glockner *et al.*, 2001), which have been suggested as a source of new miRNAs and a link between siRNA and miRNAs in humans (Piriyapongsa & Jordan, 2007). One of the miRNA genes is located within an intron. It can however not be regarded as a canonical mirtron since the splice sites do not coincide with the base of the predicted pre-miRNA stem loop. This miRNA also exhibits some other particular features which are further described in section 2.3.2.

The predicted pre-miRNA precursors are fairly long (100-200 nt) and resemble plant pre-miRNAs more than those found in animals. Furthermore, the nucleotide composition of the miRNAs was examined. For example, in plants and green algae, there is a strong bias for a 5' U, but no such preference could be observed for the miRNAs in *D. discoideum*. None of the miRNAs seem to be homologous to any known miRNAs in other organisms. We also failed to computationally identify putative homologs in *D. purpureum* although this may not be surprising considering the phylogenetic distance between the two species.

2.2.4 Differential expression of miRNAs

Many of the miRNAs were expressed at different levels in growing (0 h) cells compared to 16 h and 24 h cells. Some were almost exclusively present in the 0 h library while others were only detected in developing cells. We also found specific miRNAs expressed throughout growth and development. The relative level of miRNA and its miRNA* in the different libraries is in most cases similar. Worth mentioning is the frequent inconsistency between sequencing read counts and northern blot data we encountered during this study. These differences between the techniques have also been observed by others (Zhang *et al.*, 2010). The deep sequencing techniques all have limitations as pointed out in section 1.7.1 and northern blot data is for example strongly dependent on the labeling efficiency of the probes. In our case, ddi-mir-1177 is for example upregulated during development in northern blot analysis (Hinas *et al.*, 2007) but this is not obvious in our deep sequencing data. One problem is the lack of a reliable control, a small RNA that we can normalize the deep sequencing reads to.

2.3 Biogenesis of small RNAs in *D. discoideum* (Paper V)

2.3.1 Disruption of Dicer proteins

Disruption strains of both Dicer proteins had been constructed (in other labs) previous to this work. The two miRNAs identified prior to this study were shown to be absent in the *drnB*- strain but all other small RNAs investigated were still present. In the *drnA*- strain, all small RNAs were unaffected except for a siRNA, generated from the retrotransposon Skipper, which was upregulated (Hinas *et al.*, 2007). The notable thing was that the large population of siRNAs originating from DIRS-1 seemed to be completely unaffected in both Dicer knock-outs. This motivated us to examine both strains.

The *D. discoideum* strain where *drnA* and *drnB* had been previously disrupted is called AX2. This strain is anticipated two have one copy each of the two Dicer genes. However, another commonly used *D. discoideum* strain, AX4, has a large duplication carrying an extra copy of *drnA*. Hence, we suspected that also the AX2 strain might carry this duplication and thus two copies of *drnA*, where only one had been disrupted. Indeed, we found that the *drnA*- strain seemed to contain an undisrupted copy of the gene. Therefore a new *drnA* deletion construct was produced and a different AX2 strain which should not contain the duplication on chromosome 2 was used for transformation. The deletion construct was made so that one of the RNase III domains was removed. However, when analyzing transformants both a disrupted and a wt copy of the *drnA* gene was detected by PCR and Southern blot. Several attempts to disrupt this other copy failed and no clean *drnA*- strain was obtained. How this second copy of the gene can arise is unclear but the results suggest that Dicer A is essential to *D. discoideum.* Interestingly, the Skipper siRNA is upregulated also in this new "*drnA*-" strain.

Although the previously constructed *drnB*- strain displayed a phenotype (no miRNAs) some uncertainties regarding the disruption arouse and we decided to create our own version where one of the RNase III domains was removed. The disruption was confirmed by PCR and Southern blot. This new strain also lacked the miRNAs in agreement with data from the old *drnB*- strain. Cells missing Dicer B grow and develop as wt cells, indicating that miRNA mediated regulation is not essential in *D. discoideum* under our experimental conditions.

Notable is that we do not observe any accumulation of miRNA precursors in the *drnB*- strain. Furthermore, Dicer B has been shown to localize mainly to the nucleus (Dubin & Nellen, 2010). This suggests that Dicer B might be responsible for the initial processing of the miRNA precursor and have a role

analogous to Drosha in animals. The situation could also be similar to that in plants where one Dicer is responsible for miRNA generation by a two step processing event in the nucleus (Jones-Rhoades *et al.*, 2006).

2.3.2 Overexpression of miRNAs

Since the miRNAs appear to be expressed at very low levels under normal laboratory conditions and in some cases cannot even be detected by northern blot, an overexpression system was constructed. Overexpression would allow us to easier study miRNA biogenesis, targeting mechanisms and of course phenotypic effects when a specific miRNA is overexpressed. This kind of experimental validation has also been shown to be a very good indicator of true miRNAs (Chiang *et al.*, 2010). Four different miRNAs were expressed from extrachromosomal plasmids where the predicted miRNA precursors were cloned behind a constitutive actin promoter. These plasmids were transformed into both wt and *drnB- D. discoideum* cells. In wt cells the processed miRNA could be detected by northern blot for all constructs and the miRNA* was detected in all but one case. The expression level was approximately 30 times higher than in untransformed cells. In the *drnB-* background no miRNAs were detected supporting that the construct was processed through the "right" pathway. No obvious phenotypes were observed in the overexpression strains.

Furthermore, no other small RNAs were generated from the precursors except for in one case. When miRNA-2, which originates from an intron, was overexpressed slightly larger Dicer B independent small RNAs were detected using the miRNA probe. Small RNAs were also detected when using a probe recognizing the sequence further downstream in the hairpin. Hence, this hairpin seems to be processed through two pathways, both into a miRNA and multiple siRNAs. Considering its intron origin this could well represent a very young miRNA.

Cellular localization was determined for one of the overexpressed miRNAs. Nuclear and total RNA were analyzed by northern blot which showed that the miRNA as anticipated resides mainly in the cytoplasm (Figure 8A). Since small RNAs are methylated by Hen1 in plants, we wanted to investigate the nature of the 3' end of miRNAs in *D. discoideum*. This was done by β-elimination assay were treatment with periodate removes the 3' nucleotide from an RNA if it is not protected by a modification (e.g. methylation). The size of treated and untreated RNA can then be analyzed by northern blot. Our data show that after β-elimination the miRNAs is one nucleotide shorter which indicates that its 3' end is not modified (Figure 8B).

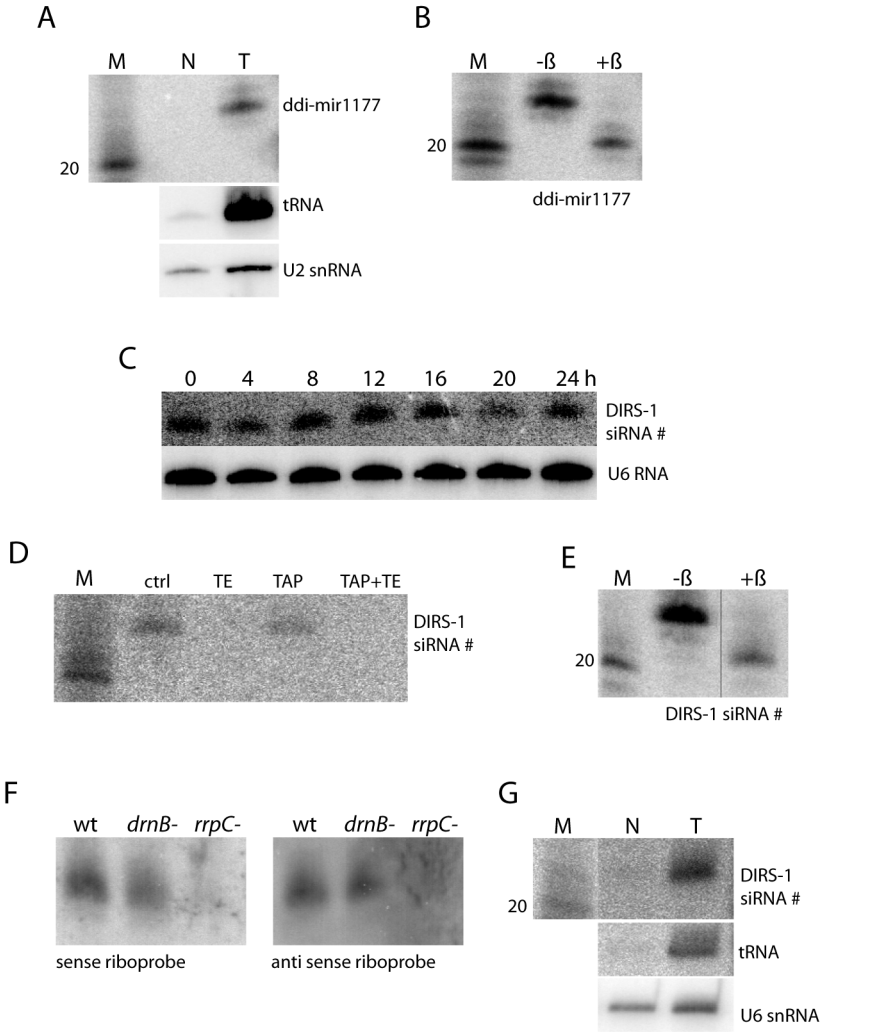

Figure 8. Northern blot data not included in manuscripts. (A) Cellular localization of miRNA 1177. (B) 3' end analysis by β-elimination assay of miRNA 1177. (C) Expression of DIRS-1 siRNA at different time points during development. (D) 5' end analysis by Terminator Exonuclease of DIRS-1 derived siRNA. (E) 3' end analysis by β-elimination assay of a DIRS-1 derived siRNA. (F) Detection of antisense and sense DIRS-1 siRNAs by 300 nt long riboprobes in wt, *drnB*- and *rrpC*- strains. (G) Cellular localization of a DIRS-1 derived siRNA. M: marker, N: nuclear RNA, T: total RNA, +/- β: RNA treated or not treated with periodate. TE: RNA treated with Terminator Exonuclease, TAP: RNA treated with Tobacco Acid Pyrophosphate. For a description of DIRS-1 siRNA #, see Figure 10.

2.4 Studying miRNA targeting mechanisms (unpublished)

To understand how the *D. discoideum* miRNAs can regulate their targets and which these targets are, two different approaches were taken. In plants, miRNAs usually bind to coding sequence of mRNA with complete or almost complete complementarity and guide Ago induced cleavage of the target at a specific site. Such putative targets were computationally predicted and 5' rapid amplification of cDNA ends (RACE) was then used to try and detect the specific cleavage site. Only a few good putative targets were predicted and in no case the correct cleavage site was identified. Thus, there is no proof that the miRNAs work in a plant like fashion.

In animals miRNAs usually target the 3' UTR of mRNAs and mediate mRNA destabilization and/or translational inhibition. To investigate if the *D. discoideum* miRNAs can regulate a target through this mechanism a reporter system was created. The reporter system is based on two extra chromosomal plasmids (Veltman *et al.*, 2009). One plasmid contains the red fluorescent protein (RFP) reporter gene behind an actin promoter fused to a 3' UTR containing miRNA target sequences. Since it is well known that the cell to cell variation in expression from extra chromosomal plasmids in *D. discoideum* is very high, a second gene expressing green fluorescent protein (GFP) is also present in the plasmid. GFP expression should not be directly affected by the miRNA and will serve as an internal control. The other plasmid is the miRNA overexpression plasmids described in section 2.3.2. Combinations of miRNAs and reporters with different 3' UTRs were transformed into both wt and *drnB*- cells. The cells were then analyzed by flow cytometry where both RFP and GFP levels were measured. The ratio of RFP to GFP was then determined for the population in presence or absence of the correct miRNA (Figure 9B). If the miRNA can cause either mRNA destabilization or inhibit translation of RFP the ratio RFP to GFP should decrease in the population. Initial experiments with two fully complementary target sequences in the 3'UTR indicated that RFP levels decreased slightly when the miRNA was overexpressed. However, when the study was expanded to include both perfect and bulged target sites for two different miRNAs and repeated several times, it was clear that the same difference in RFP expression could be observed between two identical transformations as in response to miRNA overexpression. This is of course not sufficient to draw the conclusion that miRNAs cannot target the 3'UTR of an mRNA. But at least these experiments did not contribute with any evidence in this direction.

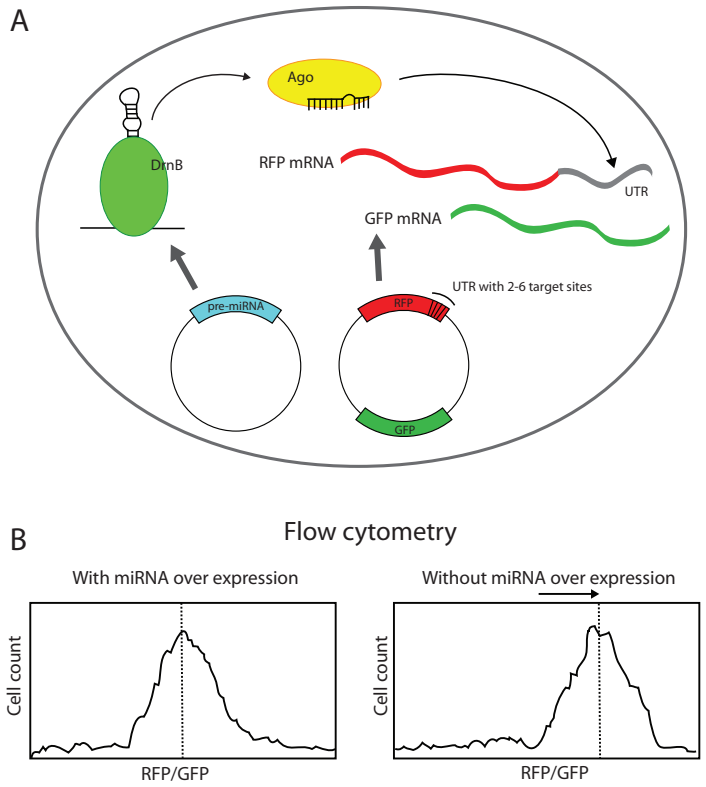

Figure 9. miRNA targeting reporter system. (A) Schematic overview of the reporter system in the cell. (B) Example of how the ratio RFP/GFP may shift in a cell population when a miRNA is absent (right) compared to overexpressed (left).

2.5 Generation of small RNAs from DIRS-1 (unpublished)

In our SOLiD library, 88% of the 21 nt RNAs are derived from DIRS-1 and covers the entire retrotransposon (Figure 10). This is in line with previous reports (Hinas *et al.*, 2007; Kuhlmann *et al.*, 2005). The large amount of transcripts could suggest some kind of amplification and generation of secondary siRNAs. This is common in some other organisms and is dependent on RdRPs. This amplification of siRNAs can be due to different species-specific pathways (see section 1.5.4). To investigate if DIRS-1 derived siRNAs are dependent on any of the RdRPs, *rrpA, B* and *C* deletion strains were created in the laboratory of C. Hammann. The level of different siRNAs in the

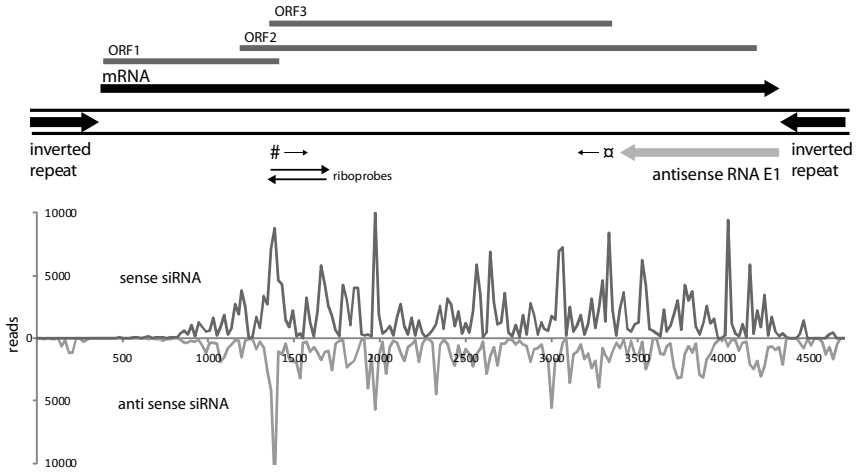

Figure 10. Distribution of sense and antisense siRNAs derived from DIRS-1. # and ¤ describes the position of the two siRNAs that were primarily investigated. Arrows indicate the direction of siRNAs and riboprobes.

deletion strains was analyzed by northern blot. This revealed that the siRNAs were dependent on RdRP C to various extents. An approximately 900 nt antisense transcript (E1) has been shown to be produced from DIRS_1 (Figure 10). siRNAs generated far downstream from this antisense transcript were almost completely dependent on RdRP C while siRNAs close to the antisense transcript was present at much more similar levels with or without RdRP C. A siRNA generated from Skipper was on the contrary significantly upregulated in the *rrpC*- strain (S. Wiegand, personal communication). Interestingly, there is also a difference in siRNA levels during development. The siRNA generated from the 3' end of the DIRS-1 mRNA (siRNA ¤) is upregulated at 16 and 24 h of development compared to growing cells (Hinas *et al.*, 2007), while a siRNA closer to the 5' end (siRNA #) is present at equal levels (Figure 8D).To investigate how these siRNAs are generated we wanted to explore the nature of their 5'-end. The SOLiD library is dependent on 5' monophosphates, but since the number of reads did not correlate with northern blot signal we were not sure if the library represent the true population or if siRNAs with other 5' ends are also present in the cell. A method which relies on the enzyme Terminator Exonuclease that degrades RNAs with 5' monophosphates but not RNA with tri-phosphates was employed. As a control we used RNAs pre-treated with Tobacco Acid Pyrophosphatase, which removes cap and other 5' structures and hence should be degraded by Terminator Exonuclease. The RdRP C dependent siRNA # was mainly used for these analyses and we could show that the siRNA must have a monophosphate 5' end which probably is generated by a

Dicer (Figure 8D). The siRNA was also analyzed by β-elimination which shows that its 3' end is not protected by modification (Figure 8E).

Using northern blot with short DNA probes it was very difficult to detect several of the siRNAs in antisense orientation to the DIRS-1 mRNA. We therefore created 300 nt long riboprobes to simultaneously detect multiple siRNAs sense or antisense to DIRS-1 mRNA. This revealed similar levels of sense and antisense siRNAs. It also showed that the siRNAs in general are dependent on RdRP C but do not require Dicer B (Figure 8F).

In conclusion it seems likely that the antisense transcript E1 serves as a starting point for dsRNA formation. RdRP C may then be recruited to produce long dsRNA which is processed by a Dicer into 21 nt RNAs. Since the DIRS-1 siRNAs analyzed are all present in *drnB*- we suspect that the "unknockable" Dicer A is responsible for their generation. Low levels of siRNAs could be detected also in the rrpC- strain. This could suggest that transcription generating low levels of the antisense RNA can be extended to cover the whole mRNA or that some additional mechanism for siRNA generation is available. Subcellular localization of the DIRS-1 siRNA # was determined by northern blot of nuclear and total RNA. This siRNA was shown to mainly reside in the cytoplasm (Figure 8G).

2.6 Spliceosomal RNAs (Paper I)

In the cDNA library from 2004 (Aspegren *et al.*, 2004), the U2 snRNA was identified in *D. discoideum*. To identify the remaining major snRNA genes (U1, U4, U5, U6) a computational approach was used which combined sequence homology, secondary structure and RNA-RNA interactions. Using these criteria, 18 loci encoding snRNAs were found and all except one were shown to be expressed. Only U6 RNA is transcribed from a single gene. The majority of genes appear in highly similar pairs separated by approximately 200-400 bp. This is similar to other small RNA genes in *D. discoideum* and other Dictyostelia which commonly occur in clusters (this study) and (Eichinger *et al.*, 2005; Aspegren *et al.*, 2004).

One interesting feature of the *D. discoideum* snRNAs uncovered here was that most of them are polyadenylated to some extent. In eukaryotes polyadenylation has mostly been described for mRNAs as a way of regulating transcript stability and translation efficiency, but there are some reports of polyadenylation of ncRNAs for example in yeast (Win *et al.*, 2006; Watanabe *et al.*, 2002). However, this is the first example of polyadenylated snRNAs where no genes involved in RNA processing have been altered.

The next unexpected finding was the unusual characteristics of some of the U2 genes. The seven U2 genes can be divided into two groups based on sequence similarity. Three genes are highly similar to each other and most closely related to U2 genes in other organisms. The other four genes are also very similar to each other but distinct from the first group in primary sequence. Furthermore they contain approximately 40 nt 5' extensions which are predicted to fold into short stem loops. In addition, the second group of U2 genes is downregulated during development and mainly located in the cytoplasm in contrast to the other snRNAs, and of course snRNAs in general.

2.7 Small ncRNA connected to host-cell interactions in *Giardia lamblia* (Paper II)

Giardia lamblia is a diplomonad which belongs to the excavata supergroup (Figure 1). *G. lamblia* is an extracellular parasite that infects humans and other mammalian hosts. It colonizes and reproduces in the small intestine but little is known about how *G. lamblia* causes disease. For example, only a few virulence factors have been identified (Ankarklev *et al.*, 2010). In an attempt to identify new virulence factors, Staffan Svärd and colleagues have investigated changes in gene expression in *G. lamblia* during interaction with differentiated Caco-2 cells (human intestinal epithelial cells) by microarrays. Expression profiles were studied after different periods of interactions with Caco-2 cells and also during stress conditions.

In addition to protein-coding genes, a number of transcripts spanning short ORFs predicted not to be transcribed were among the most up-regulated ones. It was hypothesized that these short ORFs could encode ncRNAs. To resolve this, a collaboration was initiated where we analyzed one of these putative ncRNAs. We could verify the expression of a 60 nt transcript by northern blot analysis. The RNA is upregulated after interaction with Caco-2 cells but also during stress conditions. In the absence of an established loading control for the northern blot, a probe that detects the predicted 5S rRNA was used. 5S rRNA was previously believed to be missing in *G. lamblia* but our data show the contrary. The expression of 5S rRNA is in contrast to the 60 nt ncRNA slightly reduced after interaction with host cells. The ncRNA is predicted to form a strong stem loop which resembles animal pre-miRNA. A weak signal corresponding to approximately 24 nt was also detected by northern blot, but whether this corresponds to a true small RNA is yet to be determined. We also attempted to map the 5' and 3' ends by RACE, but this was not possible probably due to the strong secondary structure of the RNA.

2.8 Multiple loss of SRP Alu-domain throughout eukaryotic evolution (Unpublished)

The SRP is responsible for localization of proteins in all kingdoms of life. However, both the nature of SRP RNA and number of proteins constituting the particle can vary between organisms. The Alu part of the SRP RNA together with SRP9/14 is responsible for translational arrest.

The variability of the Alu-domain can make it difficult to identify the SRP RNA genes. Eukaryotic SRP RNA is normally about 300 nt but for example yeast has insertions making the RNA 500-600 nt. In trypanosomes, an additional tRNA like RNA (sRNA-85) is proposed to take the function of the missing SRP14 (Lustig *et al.*, 2005; Liu *et al.*, 2003). It has been suggested that some unicellular organisms even lack the Alu-domain completely, but this has never been shown experimentally.

In collaboration with Magnus Alm Rosenblad and Staffan Svärd we wanted to use computational searches in combination with experimental approaches to investigate if indeed some organisms contain a SRP without an Alu-domain. In this study the SRP RNAs from a diverse set of protists, which all lack the SRP9/14 proteins, including *Giardia lamblia*, *Spironucleus salmonicida*, *Encephalitozon cuniculi* and *Cyanidioschyzon merolae* were investigated (Figure 11). The diplomonads *G. lamblia* and *S. salmonicida* are parasites and belong to the kingdom excavata. *E. cuniculi* is a microsporidia which are a group of intracellular parasites related to fungi (Opisthokonts). *C. merolae* is an ultra-small red algae (Archaeplastida) which lives in hot springs and have an extremely reduced genome (Nozaki *et al.*, 2007).

2.8.1 Predicting SRP RNA genes

Bioinformatic predictions of SRP RNA candidates in all four genomes were done using only the approximately 150 nt core of the S-domain of the SRP RNA. Comparative sequence analysis and putative poly-T termination signals were also used to establish the structure and sequence of the SRP RNA candidates. Predictions were then analyzed by northern blot and 5' RACE to confirm expression and determine the size and boundaries of the RNA. Using this strategy we can identify SRP RNAs either lacking the Alu part completely or containing a non-consensus version of it.

2.8.2 The missing Alu-domains

Northern blot confirmed expression of an approximately 210 nt long SRP RNA in *G. lamblia*. The 5' end could be verified by RACE and the 3' end is predicted to fold into a short stem. In *S. salmonicida* the SRP RNA is 190 nt based on northern blot data. The 5' and 3' boundaries were approximately

determined using probes mapping to genomic sequence within and outside the predicted RNA and it seems like the 3' helix found in *G. lamblia* is not present. The size of the SRP RNA in *E. cuniculi* is also around 210 nt. Probes mapping to the 5' and 3' ends confirmed a predicted 5' helix, it is possible that *E. cuniculi* possess a short 3' extension but again no Alu RNA. The S-domain of the SRP RNA in *C. merolae* was predicted to include a large insertion. This was confirmed by northern blot, which revealed a size of almost 240 nt. The boundaries of the RNA were again mapped by several northern probes but no extensions in either end were detected. In conclusion, SRP RNA candidates could be identified with high confidence in all four genomes. These are all expressed but lack the Alu part of the RNA completely (Figure 11).

Several other microsporidia genomes were also investigated and SRP RNAs, in all cases lacking identifiable Alu RNA, could be predicted. All other opisthokonts analyzed so far have SRPs with intact Alu RNA and SRP9/14 proteins. This indicates that the Alu-domain was lost early in microsporidia evolution. A similar situation seems reasonable in diplomonads (*G. lamblia* and *S. salmonicida*) as several other distant excavates (*Trichomonas vaginalis*, *Naegleria gruberi*) contain a complete Alu-domain. The loss of the Alu-domain in *C. merolae* is probably more recent since the SRP RNA of the closely related red algae *Galderia sulphuraria* has the Alu part.

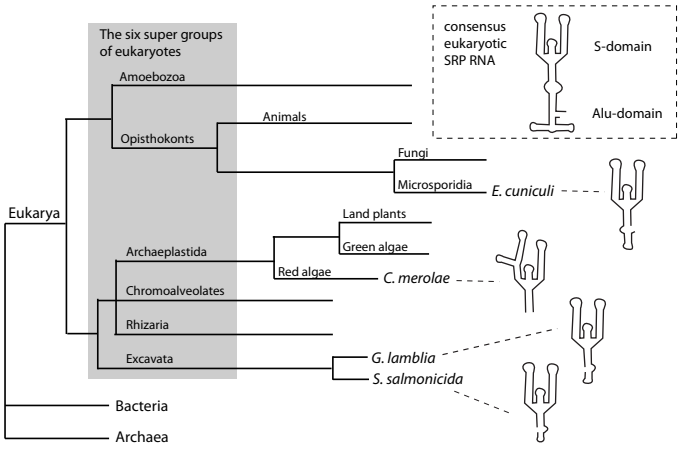

Figure 11. Schematic drawings of the SRP RNAs identified in different organisms. The consensus eukaryotic SRP RNA is shown for reference. Branch lengths in the phylogenetic tree are not to scale.

3 Discussion

3.1 Class I RNA function

The discovery of new ncRNAs has accelerated in the last decade. It is however many times not trivial to elucidate their function. This study shows that a specific class of ncRNAs, Class I RNAs is conserved and highly abundant throughout social amoebae, indicating that they have an important role in these organisms. We have used an array of different techniques to investigate their biological role but more analysis will be necessary to pinpoint their function.

3.1.1 Molecular mechanisms

The stem-forming property of Class I RNAs is under strong evolutionary pressure. The exact sequence of the stem seems to be less important but the 11 nt sequence motif is highly conserved and may serve as a binding platform for other molecules. One hypothesis is that Class I RNAs function as guides for proteins, similar to snoRNAs. In this scenario, the variable part of Class I RNAs can bind to other RNAs by complementary base pairing. Associated proteins are in this way guided to their target were they can exert their function. This is however somewhat hard to imagine considering the high sequence similarity between different Class I RNAs in the variable part between the conserved 5' and 3' ends. The antisense elements of snoRNAs in *D. discoideum* are for example much more heterogeneous (Aspegren *et al.*, 2004). We also fail to find obvious targets by computational approaches. If Class I RNAs are guide RNAs some distinct kind of target recognition must be utilized.

Another hypothesis is that Class I RNAs could act as "molecular sponges" that bind to certain proteins and regulate their activity, for example inhibit their binding to other nucleic acids. Y RNAs have for example been suggested as a repressor of Ro RNP which in its free state binds to misfolded rRNAs (O'Brien & Wolin, 1994).

The protein (CIBP), identified as an interaction partner of Class I RNAs, did unfortunately not give any hints towards the function of the complex. Apart for the two RRMs in CIBP, there is no similarity to other domains. Homologs to CIBP could be found in all Dictyostelia species investigated. By comparing the aa sequences we could see that the part connecting the RRMs is much less conserved than the RRMs themselves. This further supports that no additional functional domain is present in the protein. One hypothesis is that CIBP sequester Class I RNAs in general, inhibiting them from participating in a functional complex. Alternatively, Class I RNAs may sequester CIBP in a situation similar to Y RNAs and Ro RNP.

3.1.2 Class I RNA evolution

In this study we failed to find Class I RNAs outside of Dictyostelia. This could very well mean that these RNAs evolved in this lineage and are somehow important for the particular life cycle of social amoebae. On the other hand, it has been proven very difficult to find distant homologous ncRNA genes by *in silico* methods (Menzel *et al.*, 2009). I believe that if Class I RNAs were present (at somewhat comparable levels) in well studied model organisms like *S. cerevisiae* or *C. elegans* they should have appeared in some cDNA library. It should be noted, however, that no such analyses have been done for other amoebozoans, like *P. polycephalum*. The nematode sbRNAs are in many ways similar to Class I RNAs and could, although they do not share sequence homology, represent functional homologs.(Boria *et al.*, 2010).

It is difficult, not to say impossible, to confidently untangle the phylogenetic relationships between short ncRNAs like Class I RNAs. These genes are often evolving rapidly and in ways we cannot predict or draw information from. However, sequence information and lack of synteny between Class I genes in *D. discoideum* and *D. purpureum* suggest that the gene expansion largely took place after speciation. This is fascinating considering the large number of genes in the different species, and indicates that either the targets regulated by Class I RNAs are increasing or that very high levels of these RNAs are necessary in the cell. At least in *D. discoideum* the Class I RNAs are not completely redundant since a phenotype was observed when a single gene was disrupted.

Considering the evolutionary depth of Dictyostelia and the differences in morphology, use of chemoattractants and many other traits, Class I RNAs must be involved in some fundamental mechanism. We observe a subtle phenotype with smaller fruiting bodies following the disruption of one of the most highly expressed Class I genes in *D. discoideum*. Since acquisition of multicellularity

through aggregation is unique to Dictyostelia, it is tempting to speculate that Class I RNAs play a role in this pathway.

3.2 The RNAi pathways and associated small RNAs in *D. discoideum*

3.2.1 RNAi components

The RNAi components in *D. discoideum* exhibit some unusual features. The two Dicer proteins lack the helicase domain normally found in Dicers. Interestingly, Dicer-like helicase domains are instead found in the three RdRPs. Is it possible that complexes are formed by the Dicers and RdRPs? Some of our data supports this. In the *rrpC*- strain all Dicer B dependent miRNAs investigated are upregulated. RdRP C was also shown in this study to be responsible for generation of secondary siRNAs from DIRS-1, probably by producing long dsRNA. These siRNAs are however not dependent on Dicer B. The function of the Dicer helicase domain is not understood but it has been suggested that it could play a role in the loading of the small RNA into RISC (Jinek & Doudna, 2009) Perhaps when RdRP C is missing, miRNAs are not efficiently loaded into RISC and accumulate in the cell.

Both Dicer proteins in *D. discoideum* also lack the PAZ domain which has been shown in species ranging from *G. lamblia* to humans to be responsible for the generation of small RNAs of a fixed size. The distance between the PAZ domain which binds to the dsRNA termini and the RNase III active sites function as a ruler (Park *et al.*, 2011; MacRae *et al.*, 2007). There are other organisms with non-canonical Dicers lacking PAZ domain and it is suggested that the RNAs generated by such Dicers are more heterogeneous in size. However, a distinct mechanism of such non-canonical Dicers was recently presented for some budding yeasts. In this model, multiple Dicers bind to the dsRNA so that the distance between the proteins functions as a ruler independent of the dsRNA termini (Weinberg *et al.*, 2011). It is possible that the Dicers in *D. discoideum* could work in a similar way since our deep sequencing data suggests that both siRNAs and miRNAs are almost exclusively 21 nt. It could also be that other components interact with Dicer analogous to DGCR8 which recognizes Drosha substrates in animals (Han *et al.*, 2006).

Despite repeated efforts in creating a *drnA*- strain in *D. discoideum*, this has not been achieved and there are several mysteries surrounding this gene. Even though a complete copy of the gene clearly exists, possibly in addition to the deleted one, one phenotype has been reported for several independently created *drnA* “knock-outs” and this is the upregulation of Skipper generated siRNAs.

The Skipper siRNAs are also upregulated in the *rrpC*- strain (Hinas *et al.*, 2007) and (Stephan Wiegand, personal communication), although they are not dependent on Dicer B. Silencing of Skipper has also been shown to be disrupted in the *drnA* "knock-out" strain (Kuhlmann *et al.*, 2005), which is somewhat contradictory to elevated levels of siRNAs. The role for Dicer A thus remain unknown but it is plausible that it is involved in the generation of siRNAs and somehow essential for cell viability.

No phenotype has yet been distinguished (e.g. all endogenous small RNAs investigated remain at wt level) in *rrpA*- and *rrpB*- strains (S. Wiegand, personal communication). It is possible and has also been suggested that RdRP A and RdRP B are involved in an exogenous RNAi pathway (Martens *et al.*, 2002). The two genes are furthermore highly similar and the proteins may have redundant functions. The third class of proteins involved in RNA silencing and present in *D. discoideum* are the Argonautes. All five Agos in *D. discoideum* are PIWI-like and have the catalytic triad which is required for, but not sufficient for cleavage of a target RNA. However, next to nothing is known about the function of the *D. discoideum* Agos. Fortunately efforts are made in several labs, including ours, to study these key players of the RNAi machinery and data should hopefully start emerging shortly.

Due to the confusion surrounding several of the RNAi component knock-out strains, the Dicty/RNA community recently made a common decision regarding which strains to use and some other standardization in methodology. Efforts were also made to divide the work among the different groups. This will hopefully lead to new, exciting and reliable results in the near future.

3.2.2 DIRS-1 and its regulation by siRNAs

DIRS-1 has obviously shaped the genome of *D. discoideum* to a great extent. Fragments of this element have as mentioned been suggested to constitute the centromeres of the six chromosomes. The generation of siRNAs from this element is very extensive and represents the majority of 21-mers in our small RNA libraries. The small RNAs cover the entire element which initially was surprising since the antisense transcript E1 was reported to be only 900 nt. Our data now suggest that RdRP C is responsible for extending the dsRNA from E1 to the 5' end of the DIRS-1 mRNA. The dsRNA is then probably cut by Dicer A into siRNA duplexes (Figure 12). This is supported by the dependence on RdRP C for siRNA generation, which is more pronounced far away from E1 than close. We have investigated the cellular location for only one DIRS-1 siRNA and it is mainly cytoplasmic. If the siRNAs are responsible for chromatin maintenance at centromeres, in a situation similar to that in *Schizosaccharomyces pombe* (Volpe *et al.*, 2002), we would expect to find

them in the nucleus. It is of course possible that only a fraction of the siRNAs is necessary in the nucleus, but this suggest that at least one role is to target DIRS-1 mRNA in the cytoplasm. If the siRNAs have an important role in chromatin segregation, this could explain why Dicer A seems to be essential in *D. discoideum*.

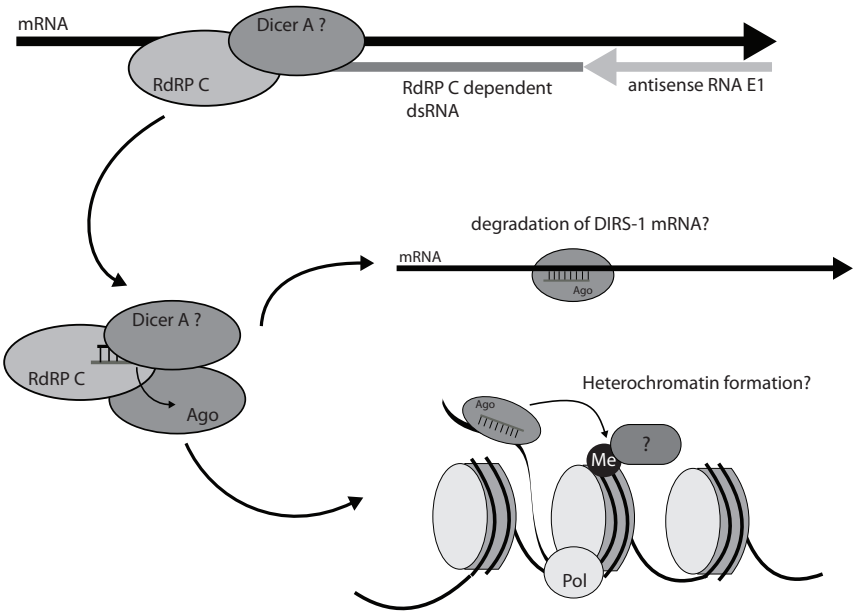

Figure 12. Model describing how siRNAs may be generated from DIRS-1 and mediate silencing.

3.2.3 Function and origin of miRNAs in *D. discoideum*

It is now beyond doubt that there are miRNAs in *D. discoideum*. The miRNAs found in this study fulfill all criteria when it comes to biogenesis. The important task now is of course to reveal the targets of these miRNAs and by what mechanism they are regulated. The miRNAs in *D. discoideum* are clearly not essential, at least not under laboratory conditions, since no phenotype is observed for the *drnB*- strain which appears to lack miRNAs completely. However, the fact that many miRNAs are expressed at different levels during development does suggest some kind of function.

We have attempted to find targets by using the knowledge from plant and animal miRNA pathways. We have identified a few, plant-like target candidates. Some of these were also tested by 5' RACE without evidence of cleavage at the predicted position. This could of course be due to that the cleavage products are too unstable for identification by 5' RACE. Still, the low

number of good such putative targets found by bioinformatics speaks against this type of targeting.

If the miRNAs find their targets in a more animal-like way, target prediction becomes very difficult. The seed sequence of several miRNAs is very AU-rich. The 3' UTRs in *D. discoideum* are also extremely AU-rich. This combination yields large numbers of potential targets. Using the reporter system described in section 2.4 we have failed to find evidence that miRNAs can regulate a target via its 3' UTR. There are however many reasons why we cannot exclude this kind of targeting. For example could the expression of RFP be too high compared to the number of miRNA, or the design of the UTR and target sites may be "wrong". It is of course also possible that the miRNA pathway in *D. discoideum* is different from both animal and plants and utilizes new rules for targeting. A hopefully useful approach to find targets is to compare mRNA levels (and optimally protein levels) on a large scale in different strains, depleted of or overexpressing specific miRNAs. If we knew the natural targets it would hopefully also be easier to determine the regulation mechanism.

None of the *D. discoideum* miRNAs share distinguishable homology to any known miRNAs in other organisms. This is however not surprising, the clearly plant-like (in terms of targeting mechanism etc.) miRNAs in the unicellular green algae *C. reinhardtii*, for example, do not share homology to any known plant miRNAs (Molnar *et al.*, 2007; Zhao *et al.*, 2007). Simple searches in the *D. purpureum* genome also fail to identify miRNA homologs.

One of the pre-miRNAs is generated from an intron and appears to be processed through two pathways, into both a miRNA and siRNAs. Multiple small RNAs from the same miRNA precursor, generated through various pathways, have been reported previously (Chellappan *et al.*, 2010; Zhang *et al.*, 2010). Whether both the Dicer B dependent miRNA and the siRNAs in our case have a function remains to be investigated. It is possible that this is a miRNA under construction. Its intronic position allows for "automatic" transcription of something that happens to fold into a stem loop which is then processed into small RNAs.

3.3 Evolution of RNAi and miRNA pathways in eukaryotes

miRNAs have now been found multiple times outside the animal and plant lineages. The presence of miRNAs in many "lower" eukaryotes is however not acknowledged by everyone. Recent examples include Park *et al*. that claims without further explanation that there is no miRNA pathway in *G. lamblia* (Park *et al.*, 2011), and Braun *et al*. that declare that they found the second

examples of miRNAs in a unicellular organism (*T. gondii*) in 2010 next to the report in *C. reinhardtii* (Braun *et al.*, 2010), neglecting the findings in for example *D. discoideum*. And perhaps it is difficult to distinguish the different RNAi pathways from each other and define exactly which small RNAs that qualifies as miRNA. Maybe miRNA-like genes are "easy" to invent in a complex genome if there is a functional RNAi machinery present? Recent work in animals show that new miRNAs genes can emerge easily at least in regions of DNA which are already transcribed like introns (Campo-Paysaa *et al.*, 2011).

Could the miRNA pathway in *D. discoideum* represent a recent *de novo* invention still under construction? Compared to other Dictyostelia species *D. discoideum* seems to have an expanded RNAi system. Homology searches by us in the *D. purpureum* genome indicates that only one Dicer and two Ago homologs is present and there is no obvious *rrpC* homolog. The situation seems to be similar in *P. pallidum, D. fasciculatum* and *A. subglobosum*. But if this corresponds to the development of a miRNA pathway in *D. discoideum* is not possible to predict.

The question of whether miRNAs are an ancient feature of eukaryotes will perhaps never be answered with confidence since the short length of these RNAs most likely limits the possibility to identify orthologs between distantly related species. Although miRNAs now have been found in several unicellular eukaryotes, it is still very possible that miRNAs have been important players in the evolution of multicellularity as previously suggested (Bartel, 2004). There has definitely been a significant expansion of miRNA genes in more complex multicellular organisms.

3.4 Multiple loss of Alu-domain in SRP RNA

In most eukaryotes efficient translocation of proteins is dependent on elongation arrest. This function is usually carried out by the SRP9/14 proteins which are associated with the Alu part of the SRP RNA. In mammals elongation arrest is essential for maintaining cellular functions and growth. Cellular membranes are depleted of proteins and protein secretion is reduced in absence of SRP with elongation arrest activity (Lakkaraju *et al.*, 2008). The exact mechanism by which elongation arrest is achieved is still unclear but a recent study of the human SRP identified a number of residues in SRP9/14 that are essential for this process (Mary *et al.*, 2010). These residues are however not highly conserved outside metazoans. This suggests that distinct mechanisms to achieve elongation arrest exist in different organisms. As we have shown in this study, the Alu-domain of SRP RNA has also been lost

several times during evolution. Either these organisms have a different way of accomplishing elongation arrest or alternative ways of compensating for the lack of it. It has also been shown that overexpression of the SRP receptor α or decreased translation rate can rescue the loss of translational arrest in mammals (Lakkaraju *et al.*, 2008).

3.5 Complexity, the non-coding genome and RNA regulation

What is a complex organism and what makes one organisms more complex than another? In what way are social amoebae more or less complex than other organisms? What roles are the non-coding genome and ncRNAs playing in the transition to multicellularity and development of complex traits? This is of course questions without straight forward answers.

More and more data are accumulating suggesting that the non-coding genome is the determinant of complexity. The ENCODE pilot project shows that transcription from exonic regions is normally similar between many cell lines while transcription from intronic and intergenic regions were much more cell type specific (Birney *et al.*, 2007). This indicates that non-coding DNA regions are functionalized in new cell types which may contribute to increased organism complexity (Alexander *et al.*, 2010). In animals, a relationship between evolution of miRNAs and the establishment of tissue identity has been pointed out. One study suggests that many of the most conserved miRNAs were originally expressed only in certain tissues, for example in brain centers or sensory tissue (Christodoulou *et al.*, 2010). Analyses of animals and yeast have revealed large amounts of more or less stable transcripts of various lengths associated with promoters in both sense and antisense direction. This apparently sloppy control of transcription is suggested to allow for unconventional transcription regulation mechanisms, for example transcription interference. However it is also suggested to constitute a pool of transcripts that could become stable and acquire new functions and thus be a key in eukaryotic evolution (Jacquier, 2009).

One theory suggests that in simpler life forms (like prokaryotes), most regulatory functions were (and are still) carried out by proteins. This theory further proposes that at some point there is a limit to how many regulatory proteins a genome can harbor. When this limit was reached in eukaryotes, the regulatory role of RNA was extended, allowing for increased complexity (Mattick, 2009a). RNA based regulation is proposed as the ideal solution for adaption to environmental changes or the transition from single to multicellularity since these mechanism are rapidly evolving (Collins & Chen, 2009).

Can we learn anything from social amoeba? Although the genomes of *D. discoideum* and *D. purpureum* differs as much as those of human an bony fish, the two species are morphologically very alike and share a very similar transcriptome under development (Parikh *et al.*, 2010). A massive change in gene expression is observed at the shift from unicellularity to multicellularity, but so far the transcription factors expected to be responsible have not been identified, despite many years of searching. Is it possible that ncRNAs are involved in the regulation of this transition?

3.6 Ancient RNAs and their evolution

Non-coding sequences can evolve more freely than protein-coding DNA due to more relaxed constraints on the primary sequence. A single insertion in a protein-coding gene can often lead to a completely useless protein, while the effect on a ncRNA usually is much less profound. There are of course exceptions. A mutation in the seed sequence of a miRNA could for example completely change its targeting preferences and no mutations have been observed in the Xist gene. However, the regulatory nature of many ncRNAs often also make them less essential and a loss of function may not be as detrimental as for a protein (Mattick, 2009b).

LECA clearly already contained many classes of ncRNAs and RNA based regulatory systems (Figure 7) and we can see how their roles have diversified across the eukaryotic tree. An RNAi system was most likely used in LECA as a defense system against invading nucleic acids like transposable elements and viruses. Throughout evolution the RNAi machinery has adapted a variety of roles in different eukaryotic lineages but is completely lost in others. In mammals for example, an adaptive immune system has evolved to fight viral infections and the role of the RNAi pathway in this process is not clear. Instead, RNAi related mechanisms are vital for countless regulatory processes involved in for example cell differentiation, organ development and embryogenesis. Another illustration of diversification is the Y RNAs that are proposed to participate in the Ro RNP which is involved in RNA quality control, but also have a role in DNA replication through independent mechanisms (Langley *et al.*, 2010). The nematode sbRNAs are homologs to Y RNAs but their function is so far unknown. There is also the example of the tRNA like component of SRP in Trypanosomes (Lustig *et al.*, 2005). snoRNAs have been acknowledged many different roles lately including acting as precursors for small RNAs with miRNA like functions in as diverse organisms as human (Ender *et al.*, 2008) and *G. lamblia* (Saraiya & Wang, 2008). The function and evolution of lncRNAs have not been studied extensively yet, but

the lack of conservation is in general obvious also in closely related species. This is implies that these genes are evolving very rapidly and may play an important role in speciation events.

In *D. discoideum* a rich variation of ncRNAs have been identified. The ncRNA population is in many aspects typical of a eukaryote but at the same time harbors particularities. The unusual U2 snRNAs with an extra 5' stem loop for example - could the 5' elongation be a first step towards a ncRNA with partly or even completely different function? Many snRNAs in addition have an atypical (cytoplasmic) localization and polyA-tails, which may suggest multiple means of regulation or even functions. Could the evolution and expansion of Class I RNAs be a determinant for the particular form of multicellularity seen among social amoeba? This is so far only an exciting speculation.

3.7 *D. discoideum* as a model for non-coding RNAs

It has for many years been proven that *D. discoideum* is in many aspects a useful model organism. Since the basic eukaryotic ncRNA population is present, the roles of these molecules can be studied during for example differentiation or host pathogen interaction. One limitation has for long been the lack of genome sequences from related species. Fortunately, the genome projects of several other Dictyostelia are finished or almost finished. *Unfortunately*, these species are very diverse, still making comparative genomics difficult in some cases. Hopefully, the ongoing sequencing of the *D. citrinum* genome, a group 4 species closely related to *D. discoideum* will solve some of these problems.

From an evolutionary standpoint *D. discoideum* represent a major branch of eukaryotes. It exemplifies a unique example of multicellularity and is at the same time one of few free-living protists with a completed sequenced genome. Thus, *D. discoideum* is an important organism, providing understanding of the ancestral eukaryote and the ancient roles of non-coding RNAs in eukaryotes.

4 Conclusion and future perspectives

During the last years, deep sequencing techniques have generated incredible amounts of data and revolutionized molecular biology. Genomes are sequenced at a rapidly increasing speed and the costs are simultaneously reduced. Transcriptomic data is also flooding the field and it is clear that large parts of the genome are transcribed. The next challenge is no doubt to understand the function of these transcripts, a task which is much less high throughput than finding them. This thesis is largely focused on several classes of ncRNAs in *D. discoideum* and the most efforts have been put into the characterization of Class I RNAs and the miRNA pathway. We have applied many techniques, both experimental and computational, but further analysis will be necessary to elucidate the function of these RNAs and the pathways they are involved in. miRNA targets will hopefully be identified through deep sequencing of mRNAs in cells lacking and overexpressing miRNAs. Further analysis of the Argonautes and other RNAi components is already underway by construction of multiple knock-out strains, localization studies and immunoprecipitation to identify interacting molecules. Class I RNAs can be studied by immunoprecipitation of CIBP. It is also possible that a large part of the Class I RNA population can be knocked down simultaneously by RNAi.
Some of the most interesting findings and points of interest of this work are listed here:

- Class I RNAs are conserved all through Dictyostelia but seem to be absent in other amoebozoans. Could this class of ncRNAs be involved in the unique form of multicellularity found in social amoeba?
- Several important clues towards the function of Class I RNAs were uncovered, including the identification of an interacting protein and a developmental phenotype upon the disruption of one Class I RNA gene. What other molecules might be associated with this complex and which pathways are the Class I RNP involved in?

- 19 miRNA genes could be identified in *D. discoideum* with high confidence. Does the *D. discoideum* miRNA pathway work in a fashion similar to that in animals or plants, or could it be so different that other rules apply?
- The small RNA population in *D. discoideum* is dominated by siRNAs generated from DIRS-1. These seem to be dependent on RdRP C through an amplification mechanism. Are these siRNAs somehow responsible for maintenance of the centromeric regions which are made up largely by DIRS-1 fragments?

I have also been involved in the characterization of a number of other ncRNAs in *D. discoideum* as well as in several diverse protists.

- *D. discoideum* has a classical repertoire of snRNA but several unexpected findings were made during the analysis of these RNAs. What is the function of the 5' extended and polyadenylated snRNAs and what are their roles in the cytoplasm?
- Several small ncRNAs are strongly upregulated during host interaction and stress in *G. lamblia*. Could they work in a way similar to any of the stress induced bacterial antisense RNAs?
- We could reveal that the Alu-domain of SRP has been lost independently in several very diverse protists. How is elongation arrest obtained in these organisms?

References

Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M.B. (2010). Annotating non-coding regions of the genome. *Nat Rev Genet* 11(8), 559-71.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G. & Tuschl, T. (2003). A uniform system for microRNA annotation. *RNA* 9(3), 277-9.

Ankarklev, J., Jerlstrom-Hultqvist, J., Ringqvist, E., Troell, K. & Svard, S.G. (2010). Behind the smile: cell biology and disease mechanisms of Giardia species. *Nat Rev Microbiol* 8(6), 413-22.

Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* 9(11), 868-82.

Aspegren, A., Hinas, A., Larsson, P., Larsson, A. & Soderbom, F. (2004). Novel non-coding RNAs in Dictyostelium discoideum and their expression during development. *Nucleic Acids Res* 32(15), 4646-56.

Avesson, L., Schumacher, H.T., Fechter, P., Romby, P., Hellman, U. & Söderbom, F. (2011). Abundant class of non-coding RNA regulates development in the social amoeba dictyostelium discoideum. *RNA Biology* 8(6).

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P. & Bartel, D.P. (2008). The impact of microRNAs on protein output. *Nature* 455(7209), 64-71.

Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Duruflé, L., Gaasterland, T., Lopez, P., Müller, M. & Philippe, H. (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proceedings of the National Academy of Sciences of the United States of America* 99(3), 1414-1419.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2), 281-97.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136(2), 215-33.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., Kuehn, M.S. et al (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146), 799-816.

Bloomfield, G., Skelton, J., Ivens, A., tanaka, Y. & Kay, R. (2010). Sex determination in the social amoeba Dictyostelium discoideum. *Science* 330, 1533 - 1536.

Boccara, M., Sarazin, A., Billoud, B., Jolly, V., Martienssen, R., Baulcombe, D. & Colot, V. (2007). New approaches for the analysis of Arabidopsis thaliana small RNAs. *Biochimie* 89(10), 1252-6.

Bompfunewerer, A.F., Flamm, C., Fried, C., Fritzsch, G., Hofacker, I.L., Lehmann, J., Missal, K., Mosig, A., Muller, B., Prohaska, S.J., Stadler, B.M., Stadler, P.F., Tanzer, A., Washietl, S. & Witwer, C. (2005). Evolutionary patterns of non-coding RNAs. *Theory Biosci* 123(4), 301-69.

Boria, I., Gruber, A., Tanzer, A., Bernhart, S., Lorenz, R., Mueller, M., Hofacker, I. & Stadler, P. (2010). Nematode sbRNAs: Homologs of Vertebrate Y RNAs. *Journal of Molecular Evolution* 70(4), 346-358.

Braun, L., Cannella, D., Ortet, P., Barakat, M., Sautel, C.F., Kieffer, S., Garin, J., Bastien, O., Voinnet, O. & Hakimi, M.A. (2010). A complex small RNA repertoire is generated by a plant/fungal-like machinery and effected by a metazoan-like Argonaute in the single-cell human parasite Toxoplasma gondii. *PLoS Pathog* 6(5), e1000920.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L. & Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science* 320(5880), 1185-90.

Burki, F., Shalchian-Tabrizi, K. & Pawlowski, J. (2008). Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. *Biology Letters* 4(4), 366-369.

Campo-Paysaa, F., Semon, M., Cameron, R.A., Peterson, K.J. & Schubert, M. (2011). microRNA complements in deuterostomes: origin and evolution of microRNAs. *Evol Dev* 13(1), 15-27.

Carthew, R.W. & Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4), 642-655.

Cerutti, H. & Casas-Mollano, J. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. *Current Genetics* 50(2), 81-99.

Chapman, E.J. & Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet* 8(11), 884-96.

Chellappan, P., Xia, J., Zhou, X., Gao, S., Zhang, X., Coutino, G., Vazquez, F., Zhang, W. & Jin, H. (2010). siRNAs from miRNA sites mediate DNA methylation of target genes. *Nucleic Acids Res* 38(20), 6883-94.

Chen, X., Quinn, A.M. & Wolin, S.L. (2000). Ro ribonucleoproteins contribute to the resistance of Deinococcus radiodurans to ultraviolet irradiation. *Genes Dev* 14(7), 777-82.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., Blelloch, R., Schroth, G.P., Nusbaum, C. & Bartel, D.P. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev* 24(10), 992-1009.

Christodoulou, F., Raible, F., Tomer, R., Simakov, O., Trachana, K., Klaus, S., Snyman, H., Hannon, G.J., Bork, P. & Arendt, D. (2010). Ancient animal microRNAs and the evolution of tissue identity. *Nature* 463(7284), 1084-8.

Christov, C.P., Gardiner, T.J., Szuts, D. & Krude, T. (2006). Functional Requirement of Noncoding Y RNAs for Human Chromosomal DNA Replication. *Mol. Cell. Biol.* 26(18), 6993-7004.

Cock, J.M., Sterck, L., Rouze, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.-M., Badger, J.H., Beszteri, B. et al (2010). The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* 465(7298), 617-621.

Collins, L.J. & Chen, X.S. (2009). Ancestral RNA: the RNA biology of the eukaryotic ancestor. *RNA Biol* 6(5), 495-502.

Costa, F.F. (2010). Non-coding RNAs: Meet thy masters. *Bioessays* 32(7), 599-608.

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227(5258), 561-3.

Czech, B. & Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet* 12(1), 19-31.

Deng, W., Zhu, X., Skogerbø, G., Zhao, Y., Fu, Z., Wang, Y., He, H., Cai, L., Sun, H., Liu, C., Li, B., Bai, B., Wang, J., Jia, D., Sun, S., He, H., Cui, Y., Wang, Y., Bu, D. & Chen, R. (2006). Organization of the Caenorhabditis elegans small non-coding transcriptome: Genomic features, biogenesis, and expression. *Genome Research* 16(1), 20-29.

Dubin, M., Fuchs, J., Graf, R., Schubert, I. & Nellen, W. (2010). Dynamics of a novel centromeric histone variant CenH3 reveals the evolutionary ancestral timing of centromere biogenesis. *Nucleic Acids Res*.

Dubin, M. & Nellen, W. (2010). A versatile set of tagged expression vectors to monitor protein localisation and function in Dictyostelium. *Gene* 465(1-2), 1-8.

Eddy, S.R. (2002). Computational genomics of noncoding RNA genes. *Cell* 109(2), 137-40.

Eichinger, L., Pachebat, J.A., Glockner, G., Rajandream, M.A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., Tunggal, B. et al (2005). The genome of the social amoeba Dictyostelium discoideum. *Nature* 435(7038), 43-57.

The ENCODE Project Consortium (2004). *Science* 306, 636.

Ender, C., Krek, A., Friedlander, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N. & Meister, G. (2008). A human snoRNA with microRNA-like functions. *Mol Cell* 32(4), 519-28.

Ennis, H.L., Dao, D.N., Pukatzki, S.U. & Kessin, R.H. (2000). Dictyostelium amoebae lacking an F-box protein form spores rather than stalk in chimeras with wild type. *Proc Natl Acad Sci U S A* 97(7), 3292-7.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M. & Izaurralde, E. (2009). Deadenylation is a widespread effect of miRNA regulation. *RNA* 15(1), 21-32.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.F., Doerks, T., Dorner, S., Bork, P., Boutros, M. & Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev* 21(20), 2558-70.

Faix, J., Kreppel, L., Shaulsky, G., Schleicher, M. & Kimmel, A.R. (2004). A rapid and efficient method to generate multiple gene disruptions in Dictyostelium discoideum using a single selectable marker and the Cre-loxP system. *Nucleic Acids Res* 32(19), e143.

Flowers, J.M., Li, S.I., Stathos, A., Saxer, G., Ostrowski, E.A., Queller, D.C., Strassmann, J.E. & Purugganan, M.D. (2010). Variation, Sex, and Social Cooperation: Molecular Population Genetics of the Social Amoeba *Dictyostelium discoideum*. *PLoS Genet* 6(7), e1001013.

Friedman, R.C., Farh, K.K., Burge, C.B. & Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19(1), 92-105.

Gesteland, R.F., Cech, T.R. & Atkins, J.F. (Eds.) (2006). *The RNA World*.

Ghildiyal, M. & Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10(2), 94-108.

Gilbert, O.M., Foster, K.R., Mehdiabadi, N.J., Strassmann, J.E. & Queller, D.C. (2007). High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proc Natl Acad Sci U S A* 104(21), 8913-7.

Glockner, G. & Heidel, A.J. (2009). Centromere sequence and dynamics in Dictyostelium discoideum. *Nucleic Acids Res* 37(6), 1809-16.

Glockner, G., Szafranski, K., Winckler, T., Dingermann, T., Quail, M.A., Cox, E., Eichinger, L., Noegel, A.A. & Rosenthal, A. (2001). The complex repeats of Dictyostelium discoideum. *Genome Res* 11(4), 585-94.

Griffiths-Jones, S., Hui, J.H., Marco, A. & Ronshaugen, M. (2011). MicroRNA evolution by arm switching. *EMBO Rep* 12(2), 172-7.

Guo, H., Ingolia, N.T., Weissman, J.S. & Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466(7308), 835-40.

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T. & Kim, V.N. (2006). Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell* 125(5), 887-901.

Heidel, A., Lawal, H., Felder, M., Schilde, C., Helps, N., Tunggal, B., Rivero, F., John, U., Schleicher, M., Eichinger, L., Platzer, M., Noegel, A., Schaap, P. & Glöckner, G. (2011). Phylogeny-wide analysis of social amoeba

genomes highlights ancient origins for complex intercellular communication. *Genome Research*.

Hinas, A., Reimegard, J., Wagner, E.G., Nellen, W., Ambros, V.R. & Soderbom, F. (2007). The small RNA repertoire of Dictyostelium discoideum and its regulation by components of the RNAi pathway. *Nucleic Acids Res* 35(20), 6714-26.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P. & Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* 17(14), 1265-72.

Huntzinger, E. & Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* 12(2), 99-110.

Izquierdo, M.A., Scheffer, G.L., Flens, M.J., Schroeijers, A.B., van der Valk, P. & Scheper, R.J. (1996). Major vault protein LRP-related multidrug resistance. *Eur J Cancer* 32A(6), 979-84.

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet* 10(12), 833-844.

Jinek, M. & Doudna, J.A. (2009). A three-dimensional view of the molecular machinery of RNA interference. *Nature* 457(7228), 405-12.

Jones-Rhoades, M.W., Bartel, D.P. & Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. *Annu Rev Plant Biol* 57, 19-53.

Katoh, M., Chen, G., Roberge, E., Shaulsky, G. & Kuspa, A. (2007). Developmental commitment in Dictyostelium discoideum. *Eukaryot Cell* 6(11), 2038-45.

Kessin, R. (2001). *Dictyostelium*: Cambridge University Press, Cambridge.

King, J.S. & Insall, R.H. (2009). Chemotaxis: finding the way forward with Dictyostelium. *Trends Cell Biol* 19(10), 523-30.

Kozomara, A. & Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* 39(suppl 1), D152-D157.

Kuhlmann, M., Borisova, B.E., Kaller, M., Larsson, P., Stach, D., Na, J., Eichinger, L., Lyko, F., Ambros, V., Soderbom, F., Hammann, C. & Nellen, W. (2005). Silencing of retrotransposons in Dictyostelium by DNA methylation and RNAi. *Nucleic Acids Res* 33(19), 6405-17.

Kurland, C.G. (2010). The RNA dreamtime. *Bioessays* 32(10), 866-871.

Kuspa, A. & Loomis, W.F. (1992). Tagging developmental genes in Dictyostelium by restriction enzyme-mediated integration of plasmid DNA. *Proceedings of the National Academy of Sciences* 89(18), 8803-8807.

Lakkaraju, A.K.K., Mary, C., Scherrer, A., Johnson, A.E. & Strub, K. (2008). SRP Keeps Polypeptides Translocation-Competent by Slowing Translation to Match Limiting ER-Targeting Sites. *Cell* 133(3), 440-451.

Langley, A.R., Chambers, H., Christov, C.P. & Krude, T. (2010). Ribonucleoprotein particles containing non-coding Y RNAs, Ro60, La

and nucleolin are not required for Y RNA function in DNA replication. *PLoS One* 5(10), e13673.

Lee, J.T. (2010). The X as Model for RNA's Niche in Epigenomic Regulation. *Cold Spring Harbor Perspectives in Biology* 2(9).

Lee, R.C., Feinbaum, R.L. & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75(5), 843-54.

Lerner, M.R., Boyle, J.A., Hardin, J.A. & Steitz, J.A. (1981). Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science* 211(4480), 400-2.

Levine, M. & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* 424(6945), 147-151.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. (2003). Prediction of mammalian microRNA targets. *Cell* 115(7), 787-98.

Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr Biol* 15(16), 1501-7.

Li, S.I. & Purugganan, M.D. (2011). The cooperative amoeba: Dictyostelium as a model for social evolution. *Trends Genet* 27(2), 48-54.

Linsen, S.E.V., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R.K., Fritz, B., Wyman, S.K., de Bruijn, E., Voest, E.E., Kuersten, S., Tewari, M. & Cuppen, E. (2009). Limitations and possibilities of small RNA digital gene expression profiling. *Nat Meth* 6(7), 474-476.

Liu, L., Ben-Shlomo, H., Xu, Y.X., Stern, M.Z., Goncharov, I., Zhang, Y. & Michaeli, S. (2003). The trypanosomatid signal recognition particle consists of two RNA molecules, a 7SL RNA homologue and a novel tRNA-like molecule. *J Biol Chem* 278(20), 18271-80.

Llave, C., Xie, Z., Kasschau, K.D. & Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* 297(5589), 2053-6.

Lustig, Y., Goldshmidt, H., Uliel, S. & Michaeli, S. (2005). The Trypanosoma brucei signal recognition particle lacks the Alu-domain-binding proteins: purification and functional analysis of its binding proteins by RNAi. *J Cell Sci* 118(Pt 19), 4551-62.

MacRae, I.J. & Doudna, J.A. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr Opin Struct Biol* 17(1), 138-45.

MacRae, I.J., Zhou, K. & Doudna, J.A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* 14(10), 934-40.

Maida, Y. & Masutomi, K. (2011). RNA-dependent RNA polymerases in RNA silencing. *Biol Chem* 392(4), 299-304.

Martens, H., Novotny, J., Oberstrass, J., Steck, T.L., Postlethwait, P. & Nellen, W. (2002). RNAi in Dictyostelium: the role of RNA-directed RNA polymerases and double-stranded RNase. *Mol Biol Cell* 13(2), 445-53.

Mary, C., Scherrer, A., Huck, L., Lakkaraju, A.K.K., Thomas, Y., Johnson, A.E. & Strub, K. (2010). Residues in SRP9/14 essential for elongation arrest activity of the signal recognition particle define a positively charged functional domain on one side of the protein. *RNA* 16(5), 969-979.

Mattick, J.S. (2004). RNA regulation: a new genetics? *Nat Rev Genet* 5(4), 316-23.

Mattick, J.S. (2009a). Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Ann N Y Acad Sci* 1178, 29-46.

Mattick, J.S. (2009b). The genetic signatures of noncoding RNAs. *PLoS Genet* 5(4), e1000459.

Mattick, J.S. (2011). The central role of RNA in human development and cognition. *FEBS Lett* 585(11), 1600-16.

Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R. & Tuschl, T. (2005). Identification of novel argonaute-associated proteins. *Curr Biol* 15(23), 2149-55.

Menzel, P., Gorodkin, J. & Stadler, P.F. (2009). The tedious task of finding homologous noncoding RNA genes. *RNA* 15(12), 2075-2082.

Mercer, T.R., Dinger, M.E. & Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10(3), 155-9.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J. et al. *Plant Cell* 20(12), 3186-90.

Millar, A.A. & Waterhouse, P.M. (2005). Plant and animal microRNAs: similarities and differences. *Funct Integr Genomics* 5(3), 129-35.

Miyoshi, K., Miyoshi, T. & Siomi, H. (2010). Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. *Molecular Genetics and Genomics* 284(2), 95-103.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C. & Baulcombe, D.C. (2007). miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. *Nature* 447(7148), 1126-1129.

Mosig, A., Sameith, K. & Stadler, P. (2006). Fragrep: an efficient search tool for fragmented patterns in genomic sequences. *Genomics Proteomics Bioinformatics* 4(1), 56-60.

Nawrocki, E.P., Kolbe, D.L. & Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25(10), 1335-1337.

Nellen, W., Hildebrandt, M., Mahal, B., Mohrle, A., Kroger, P., Maniak, M., Oberhauser, R. & Sadiq, M. (1992). Mechanisms of gene regulation by endogenous and artificially introduced antisense RNA. *Biochem Soc Trans* 20(4), 750-4.

Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T., Takio, S. et al (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. *BMC Biology* 5(1), 28.

O'Brien, C.A. & Wolin, S.L. (1994). A possible role for the 60-kD Ro autoantigen in a discard pathway for defective 5S rRNA precursors. *Genes & Development* 8(23), 2891-2903.

Ohno, S. (1972). So much "junk" in our DNA. *Brookhaven symposia in biology* 23, 366-370.

Ozsolak, F. & Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12(2), 87-98.

Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M. & Milos, P.M. (2009). Direct RNA sequencing. *Nature* 461(7265), 814-818.

Parikh, A., Miranda, E., Katoh-Kurasawa, M., Fuller, D., Rot, G., Zagar, L., Curk, T., Sucgang, R., Chen, R., Zupan, B., Loomis, W., Kuspa, A. & Shaulsky, G. (2010). Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biology* 11, R35.

Park, J.-E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J. & Kim, V.N. (2011). Dicer recognizes the 5[prime] end of RNA for efficient and accurate processing. *Nature* 475(7355), 201-205.

Piriyapongsa, J. & Jordan, I.K. (2007). A Family of Human MicroRNA Genes from Miniature Inverted-Repeat Transposable Elements. *PLoS One* 2(2), e203.

Piriyapongsa, J., Marino-Ramirez, L. & Jordan, I.K. (2007). Origin and evolution of human microRNAs from transposable elements. *Genetics* 176(2), 1323-37.

Pool, M.R. (2005). Signal recognition particles in chloroplasts, bacteria, yeast and mammals (review). *Mol Membr Biol* 22(1-2), 3-15.

Poole, A.M. (2004). Is all that junk really regulatory RNA? *Nat Rev Genet* 5(4).

Reichow, S.L., Hamma, T., Ferré-D'Amaré, A.R. & Varani, G. (2007). The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Research* 35(5), 1452-1464.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. & Bartel, D.P. (2002). MicroRNAs in plants. *Genes Dev* 16(13), 1616-26.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. & Bartel, D.P. (2002). Prediction of Plant MicroRNA Targets. *Cell* 110(4), 513-520.

Roger, A.J. & Simpson, A.G.B. (2009). Evolution: Revisiting the Root of the Eukaryote Tree. *Current Biology* 19(4), R165-R167.

Rokas, A. (2008). The Origins of Multicellularity and the Early History of the Genetic Toolkit For Animal Development. *Annual Review of Genetics* 42(1), 235-251.

Romeralo, M., Cavender, J., Landolt, J., Stephenson, S. & Baldauf, S. (2011). An expanded phylogeny of social amoebas (Dictyostelia) shows increasing diversity and new morphological patterns. *BMC Evolutionary Biology* 11(1), 84.

Rosen, E., Sivertsen, A. & Firtel, R.A. (1983). An unusual transposon encoding heat shock inducible and developmentally regulated transcripts in Dictyostelium. *Cell* 35(1), 243-51.

Sanger, F. & Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94(3), 441-8.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74(12), 5463-7.

Santorelli, L.A., Thompson, C.R., Villegas, E., Svetz, J., Dinh, C., Parikh, A., Sucgang, R., Kuspa, A., Strassmann, J.E., Queller, D.C. & Shaulsky, G. (2008). Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. *Nature* 451(7182), 1107-10.

Saraiya, A.A. & Wang, C.C. (2008). snoRNA, a novel precursor of microRNA in Giardia lamblia. *PLoS Pathog* 4(11), e1000224.

Schaap, P. (2007). Evolution of size and pattern in the social amoebas. *Bioessays* 29, 635 - 644.

Schaap, P., Winckler, T., Nelson, M., Alvarez-Curto, E., Elgie, B., Hagiwara, H., Cavender, J., Milano-Curto, A., Rozen, D.E., Dingermann, T., Mutzel, R. & Baldauf, S.L. (2006). Molecular phylogeny and evolution of morphology in the social amoebas. *Science* 314(5799), 661-3.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R. & Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455(7209), 58-63.

Sewer, A., Paul, N., Landgraf, P., Aravin, A., Pfeffer, S., Brownstein, M.J., Tuschl, T., van Nimwegen, E. & Zavolan, M. (2005). Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics* 6, 267.

Shabalina, S.A. & Koonin, E.V. (2008). Origins and evolution of eukaryotic RNA interference. *Trends in Ecology & Evolution* 23(10), 578-587.

Sharp, P.A. (2009). The Centrality of RNA. *Cell* 136(4), 577-580.

Simpson, A.G.B. & Roger, A.J. (2004). The real [`]kingdoms' of eukaryotes. *Current Biology* 14(17), R693-R696.

Smalheiser, N.R. & Torvik, V.I. (2005). Mammalian microRNAs derived from genomic repeats. *Trends Genet* 21(6), 322-6.

Song, J., Xu, Q., Olsen, R., Loomis, W.F., Shaulsky, G., Kuspa, A. & Sucgang, R. (2005). Comparing the *Dictyostelium* and *Entamoeba* Genomes Reveals an Ancient Split in the Conosa Lineage. *PLoS Comput Biol* 1(7), e71.

Stadler, P.F., Chen, J.J., Hackermuller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretzschmar, A.K., Mosig, A., Prohaska, S.J., Qi, X., Schutt, K. & Ullmann, K. (2009). Evolution of vault RNAs. *Mol Biol Evol* 26(9), 1975-91.

Steinert, M. & Heuner, K. (2005). Dictyostelium as host model for pathogenesis. *Cell Microbiol* 7(3), 307-14.

Taft, R.J., Pheasant, M. & Mattick, J.S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 29(3), 288-99.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V. & Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313(5785), 320-4.

Wahl, M.C., Will, C.L. & Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136(4), 701-18.

Valadkhan, S. (2010). Role of the snRNAs in spliceosomal active site. *RNA Biology* 7(3), 345-353.

Walter, P. & Blobel, G. (1980). Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. *Proc Natl Acad Sci U S A* 77(12), 7112-6.

Walter, P. & Blobel, G. (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* 299(5885), 691-8.

Van Horn, D.J., Eisenberg, D., O'Brien, C.A. & Wolin, S.L. (1995). Caenorhabditis elegans embryos contain only one major species of Ro RNP. *RNA* 1(3), 293-303.

Watanabe, T., Miyashita, K., Saito, T.T., Nabeshima, K. & Nojima, H. (2002). Abundant poly(A)-bearing RNAs that lack open reading frames in Schizosaccharomyces pombe. *DNA Res* 9(6), 209-15.

Watkins, R.F. & Gray, M.W. (2008). Sampling Gene Diversity Across the Supergroup Amoebozoa: Large EST Data Sets from Acanthamoeba castellanii, Hartmannella vermiformis, Physarum polycephalum, Hyperamoeba dachnaya and Hyperamoeba sp. *Protist* 159(2), 269-281.

Weinberg, David E., Nakanishi, K., Patel, Dinshaw J. & Bartel, David P. (2011). The Inside-Out Mechanism of Dicers from Budding Yeasts. *Cell* 146(2), 262-276.

Veltman, D.M., Akar, G., Bosgraaf, L. & Van Haastert, P.J. (2009). A new set of small, extrachromosomal expression vectors for Dictyostelium discoideum. *Plasmid* 61(2), 110-8.

Williams, J.G. (2010). Dictyostelium finds new roles to model. *Genetics* 185(3), 717-26.

Williamson, B. (1977). DNA insertions and gene structure. *Nature* 270(5635), 295-297.

Wilusz, J.E., Sunwoo, H. & Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23(13), 1494-504.

Win, T.Z., Draper, S., Read, R.L., Pearce, J., Norbury, C.J. & Wang, S.W. (2006). Requirement of fission yeast Cid14 in polyadenylation of rRNAs. *Mol Cell Biol* 26(5), 1710-21.

Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences* 74(11), 5088-5090.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136(4), 669-87.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S. & Martienssen, R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science* 297(5588), 1833-1837.

Yoshida, H., Kumimoto, H. & Okamoto, K. (1994). dutA RNA functions as an untranslatable RNA in the development of Dictyostelium discoideum. *Nucleic Acids Research* 22(1), 41-46.

Zhang, C., Frias, M.A., Mele, A., Ruggiu, M., Eom, T., Marney, C.B., Wang, H., Licatalosi, D.D., Fak, J.J. & Darnell, R.B. (2010). Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* 329(5990), 439-43.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G.J., Wang, X.-J. & Qi, Y. (2007). A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. *Genes & Development* 21(10), 1190-1203.

Acknowledgements

A PhD is of course all about science, and there are some people who definitely contributed scientifically to this thesis. My thanks go to:
My supervisor **Fredrik Söderbom** for accepting me in the lab and for giving me interesting and challenging projects. Thank you for all the enthusiasm, support and encouragement.
Co-supervisor **Gerhart Wagner** for the many discussions and suggestions on all of my projects.
Andrea Hinas for invaluable introduction to the lab and involving me in the snRNA project.
Johan Reimegård for collaboration on the deep sequencing project.
Heiko Schumacher for contributions to the Class I project.
Pontus Larsson for bioinformatic support on the Class I project.
Staffan Svärd for involving me in several collaborations.
Magnus Alm Rosenblad for collaboration on the SRP project.
Emma Ringqvist for involving me in the Giardia project.
Christian Hammann and **Stephan Wiegand** for supplying RdRP deletion strains and helpful discussions.
Pascale Romby for all the help during my stay in Strasbourg.

But really, I am not so much about being strictly scientific and these years have been about so much more. So here it goes, the love bombing:

Fredrik, thank you for being more than a great supervisor! For trusting and believing in me, for all the good times at conferences and beer clubs and New Year eves! **Andrea**, you gave me pretty big shoes to fill… I admire your drive and determination so much. But more importantly you are a really good friend. I have enjoyed every lunch, fika, dinner and party and hope there are many more to come! **Åsa**, I am so glad we shared office and lab agonies the last couple of years. You are such a good person, I hope it rubbed off a bit on me!

All the past members of the lab: **Pontus H, Fredrik A, Björn, Louise A, Monika, Jonas, Anna Å, Ilona and Naresh**. It's been a pleasure!

Everyone else at Molbio SLU and X-ray during these years:

Christofer B, for being such a great friend and bringing out the most sarcastic me. All the beer (-buying, -carrying, -clubbing and -drinking), after work burgers, dinners and parties. For looking after me also when I don't have my best days. Shanghai will be great! **Gelareh**, thanks for sharing ups and downs in the lab. For always taking time and having such a big heart. I hope I will have the privilege of hearing your laugh many times in the future too! **Nisse** and **Annette**, for all the magical skating tours and climbing excursions (yes, it was fun most of the times ☺), BBQs, mojitos and Irish coffees. Thank you for being really good friends! **Martin W, Louise, Urszula, Rosie** and **Malin** for making me feel so welcome in the lab and for all the fun beer clubs, dinners and parties. **Al**, thanks for taking me on my first and only ESRF trip! **Jenny** for surfing, climbing and skating company. Both of you for being wonderful and showing me Oslo, I wish I could see you more often! **Gösta**, thank you for all the good times. Especially the days and nights sailing on Plenty, for taking me skating, skiing and climbing. For challenging me and believing in me. You taught me a lot and I am so grateful you are my friend. **Calle, Mange** and **Jakob** for letting me join some of the beer/rock evenings. **Anna J** and **Yang** for being sötnosar, AW and making the best Team Wash & Shine with me and Christofer. **Agata** for excellent co-Beer Queen time! **Micke G** for beer drinking company. **Karin V** for your always friendly face, pottery coordinating and birthday cakes. **Margareta** for keeping order in the lab and **Christer** for keeping all lab equipment and my computer in order. **Jens, Martin, Martin, Martin, Sinisa** and **Stefan** for being excellent beer drinkers and the amusing fika discussions. **Lars B** for always being extremely entertaining and fun, **Henrik K** for finish lessons, fun discussions about everything and nothing, drama skills and high AW attendance, **Christoffer L (Sixten)** for generously letting us in on you breakfast preferences and hanging out with the old people.

All of you who made and make X-ray a great place:

Emma, Lena, Alina, Anna L, Adrian, Daniel E, Henrik I, Patrik, Jonas, Sara, Ana Laura, Sanjee, EvaLena, Cha-San, Anatoly, Henrik B, Yafei, Li, Majid, Miao, Christofer L, Frida, Mats, Henrik H, Nils, Jerry, Ulla, Elleonor, Hasse, Stefan, Anton, Dee, Inger, Janos, Sherry, Alwyn, Mark, Maria, Avinash, Nina, Tex, Torsten, Lars L, Erling, Johan, Göran, Jasmine, Alexandra, Erik M, Daniel L and the rest, you know who you are…

The past and present members of the Wagner group, where I always felt at home, especially:

Gerhart for in addition to being a great researcher also are so relaxed and fun. For hanging out at conferences, giving good advice and playing the guitar. **Johan R**, I really enjoyed our collaboration, you are great at what you do and amazing at seeing the big picture. Thank you more for being a fantastic friend, for saving me from Greek guys, for sports-watching, tear-drying, music-sharing, pole-dancing in Berlin U-bahn and being the best partylok ever! **Erik H**, for discussions about science and life. For being a good friend and spending fun time Berlin, it was a fantastic last night (why did everyone stare at us in the subway) and unforgettable morning after. And let's stay away from skaldjursbuffé from now on… **Nadja**, thank you for all the talks about love and life and cheering me up when I really needed it. I had the best time in Berlin with you and we have to go to BERGHAIN soon!! **Cia**, for the many chats in the lab and great company in Chicago. **Klas**, for not yelling at me in the beginning (and not since ☺). It is always good to see you!

All the other past and present people at ICM:

Pontus, for the collaborations, but more importantly for rubbing off some nerdiness over lunches and coffees and for super company in Chicago, Cambridge and Berlin. I am glad you are coming home! **Karin T**, thank you for all the concerts, parties and dinners (I definitely owe you one), for making me feel so spontaneous and amazing in general! **Niklas** for fun company in Chicago, CL and good RNA advice. **Shi**, for always being so friendly and teaching me about real Chinese food. **Ulrika**, thanks for letting me sneak in at the ICM/micro parties ☺. **Jon**, for being so cool and knowing everything. **Johan A**, for being cool and fun and having outstanding acting skills! **Magnus L** for spinning company.
Staffan, Emma, Disa, Marie, Hava, Henrik T, Sonchita, Bhupi, Aurelie, Maaike, Amanda, Magnus J, Magnus L, Per, Nicole, Greta and everyone else.

Everyone at CNRS in Strasbourg, especially **Pascale**, **Tom** and **Pierre**, for helping me in the lab and showing me around.
Micke C for fun collaboration on dicty RNRs.

Some "outsiders" also deserve at little love:

Sanna, thank you for being there longer than anyone else, for being there always when I need it. We have shared so many laughs and tears. It is great to have a real bff and I am so thankful it is you! **Tina**, I think you are at least a little bit nerdy like me when it comes to science, thanks for all the good times through the years! **Anna, Kim** and **Maria** and **Robin,** thanks for staying fun

and fabulous! **Dimitri**, for sharing great years of studies and parties and good times in Tallinn and Barcelona. **Lisa**, for all these years of studies, basketball, Snerikes, dinners and parties. For sharing a special year with me at USF and always being an extraordinary friend!

My fantastic family who I can always count on and love so much:

Marie och **Thomas**, tack för att ni tagit hand om er bortskämda lillasyster! Ni är verkligen de bästa syskon jag kan tänka mig, jag är så glad att jag har er. **Matti**, du har varit med så länge jag kan minnas och jag litar på dig som en bror. **Eva**, du är ett så bra ”nytt” tillskott till familjen! Och så klart mina fantastiska syskonbarn **Tilda** (jag vill gärna ha en bra anledning till att se judon live under OS 2016 ☺), **Teo** och **Jonathan**! **Farfar**, tack för alla fina minnen. Några pussar till världens sötaste schnauzer **Kajsa,** för mys och uppmuntrande yl under avhandlingsskrivandet.

Till slut vill jag tacka **Mamma** och **Pappa** för att ni alltid tror på mig, stöttar mig och finns där. Ni har gett mig möjligheten att göra det jag vill och ni inspirerar mig att vara den bästa jag kan. Tack inte minst för att jag fått ockupera ert uterum och skriva största delen av avhandlingen där. Jag älskar er mycket och alltid.

Lotta